

University of Kentucky

UKnowledge

Graduate Center for Nutritional Sciences
Faculty Patents

Nutritional Sciences

5-28-2002

Genes Encoding Several Poly (ADP-Ribose) Glycohydrolase (PARG) Enzymes, the Proteins and Fragments Thereof, and Antibodies Immnoreactive Therewith

Myron Jacobson
University of Kentucky

Elaine L. Jacobson
University of Kentucky

Jean-Christoph Amé

Winston Lin
University of Kentucky, winston.lin@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/nutrisci_patents



Part of the [Medical Pharmacology Commons](#)

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Jacobson, Myron; Jacobson, Elaine L.; Amé, Jean-Christoph; and Lin, Winston, "Genes Encoding Several Poly (ADP-Ribose) Glycohydrolase (PARG) Enzymes, the Proteins and Fragments Thereof, and Antibodies Immnoreactive Therewith" (2002). *Graduate Center for Nutritional Sciences Faculty Patents*. 21.
https://uknowledge.uky.edu/nutrisci_patents/21

This Patent is brought to you for free and open access by the Nutritional Sciences at UKnowledge. It has been accepted for inclusion in Graduate Center for Nutritional Sciences Faculty Patents by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.



US006395543B1

(12) **United States Patent**
Jacobson et al.

(10) **Patent No.:** **US 6,395,543 B1**
(45) **Date of Patent:** **May 28, 2002**

(54) **GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZYMES, THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE THEREWITH**

(75) Inventors: **Myron K. Jacobson; Elaine L. Jacobson**, both of Lexington, KY (US); **Jean-Christophe Amé**, Obernai (FR); **Winston Lin**, Lexington, KY (US)

(73) Assignee: **University of Kentucky Research Foundation**, Lexington, KY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/511,507**

(22) Filed: **Feb. 23, 2000**

Related U.S. Application Data

(62) Division of application No. 09/302,812, filed on Apr. 30, 1999.

(60) Provisional application No. 60/083,768, filed on May 1, 1998.

(51) **Int. Cl.**⁷ **C12Q 1/68**; C07H 21/04; C12N 15/85; C12N 15/86

(52) **U.S. Cl.** **435/375**; 435/325; 435/6; 435/91.1; 536/23.1; 536/23.5; 536/24.3; 536/24.33; 536/24.31

(58) **Field of Search** 435/6, 91.1, 91.3, 435/325, 375; 536/23.1, 23.2, 23.5, 24.5, 24.31, 24.33, 24.3

(56) **References Cited**
PUBLICATIONS

Lin et al. "Isolation and characterization of the cDNA encoding Bovein poly(ADP-ribose) glycohydrolase" JBC, vol. 272 No. 18, pp. 11895-11901, May 1997.*

* cited by examiner

Primary Examiner—John S. Brusca

Assistant Examiner—Karen A Lacourciere

(74) *Attorney, Agent, or Firm*—Fulbright & Jaworski, LLP

(57) **ABSTRACT**

The isolation and characterization of cDNAs encoding poly (ADP-ribose) glycohydrolase (PARG) enzymes and the amino acid sequences of PARGs from several species are described. PARG is involved in the cellular response to DNA damage and its proper function is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. Expression vectors containing the cDNAs and cells transformed with the vectors are described. Probes and primers that hybridize with the cDNAs are described. Expression of the cDNA in *E. coli* results in an enzymatically active protein of about 111 kDa and an active fragment of about 59 kDa. Methods for inhibiting PARG expression or overexpressing PARG in a subject for therapeutic benefit are described. Exemplary of PARG inhibitors are anti-sense oligonucleotides. The invention has implications for treatment of neoplastic disorder, heart attack, stroke, and neurodegenerative diseases. Methods for detecting a mutant PARG allele are also described. Antibodies immunoreactive with PARGs and fragments thereof are described.

18 Claims, 18 Drawing Sheets

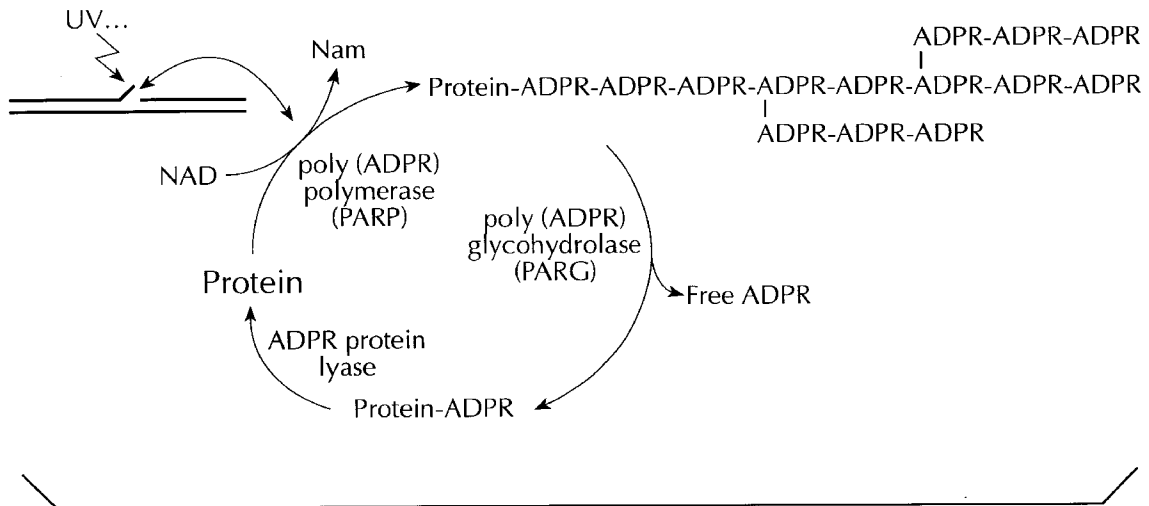


FIG. 1

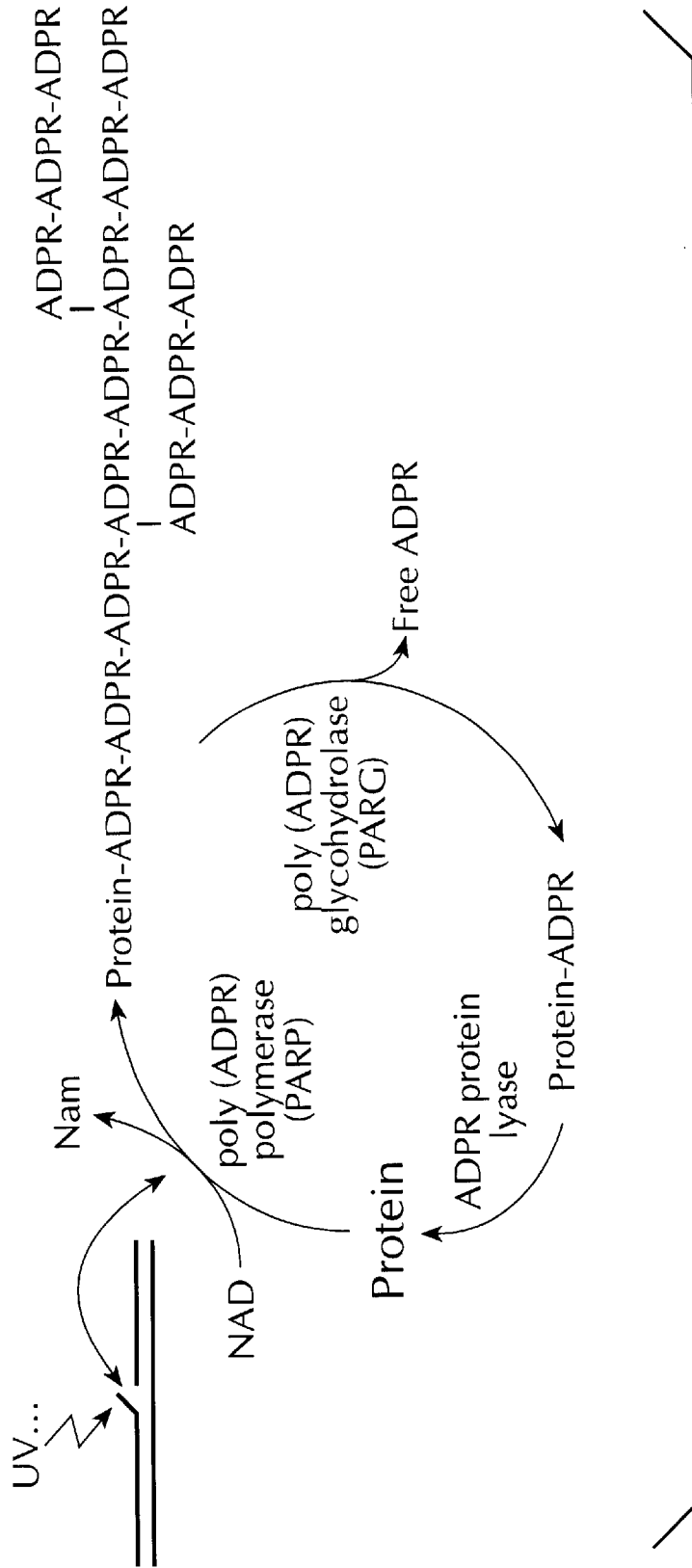


FIG. 2

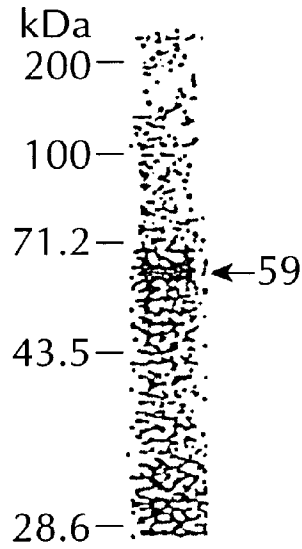


FIG. 4A

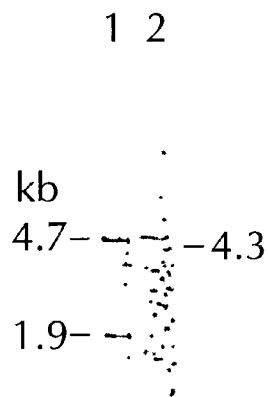


FIG. 4B

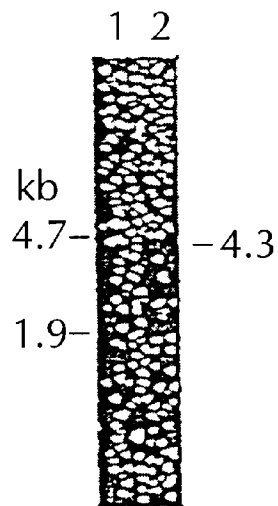


FIG. 3

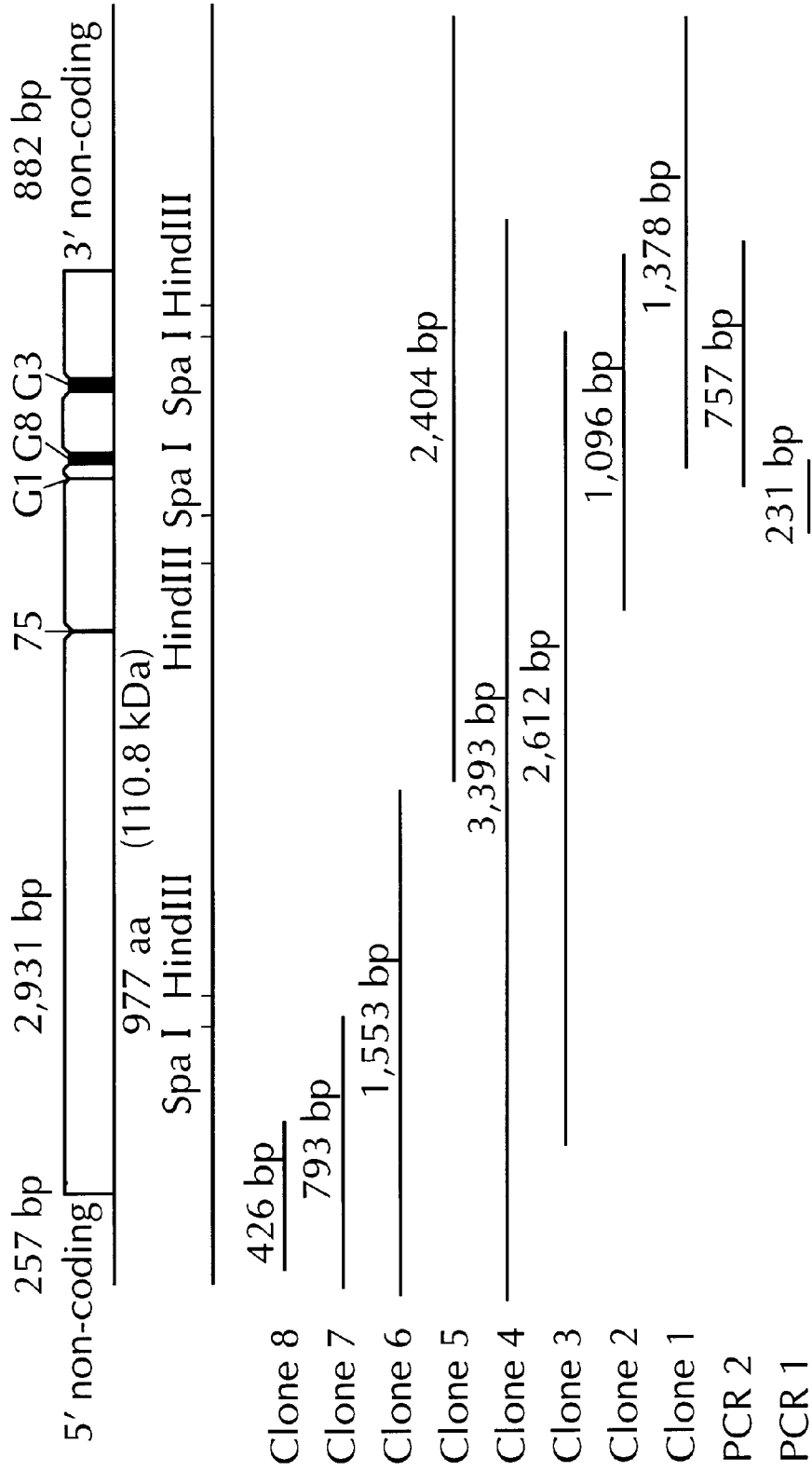


FIG. 5

bPARG (422) ED...KRKEQCEMKHQ RTE..RKIPKYIPPH SEQ ID NO: 19
hPARG (421) ED...RRKEQWETKHQ RTE..RKIPKYVPPH SEQ ID NO: 20
mPARG (413) ED...RRKEQCEVRHQ RTE..RKIPKYIPP N SEQ ID NO: 21
CePARG (29) HQVPTMKRRKLTEHGNTTESLLLKEDPEEPKS SEQ ID NO: 22
hPARP (205) EG...KRKGD.EVDG.VDEVAKKKSKEKDK SEQ ID NO: 23
mPARP (205) EG...KRKGD.EVDG.TDEVAKKKS RKEKDK SEQ ID NO: 24
bPARP (208) EG...KRKGD.EVDG.IDEVAKKKSKEKDK SEQ ID NO: 25
aPARP (205) EG...KRKGE.EVDG..NVVAKKKS RKEKEK SEQ 11 NO: 26
XIPARP (204) EG...KRKAD.EVDG.HSAARKKKIKKEKEK SEQ ID NO: 27
DmPARP (202) EELPDTKRAKM.ELSDTNEEGEKKQR..... SEQ ID NO: 28
SpPARP (205) EGVSSAKKAKI.EKIDEEDAASIKELTEKIKK SEQ ID NO: 29

FIG. 6

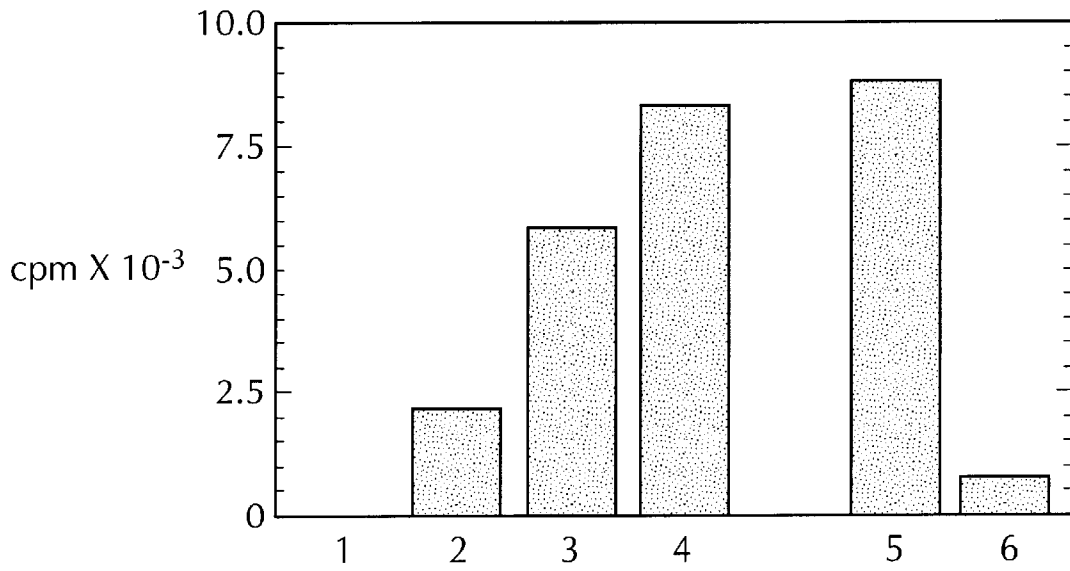


FIG. 9

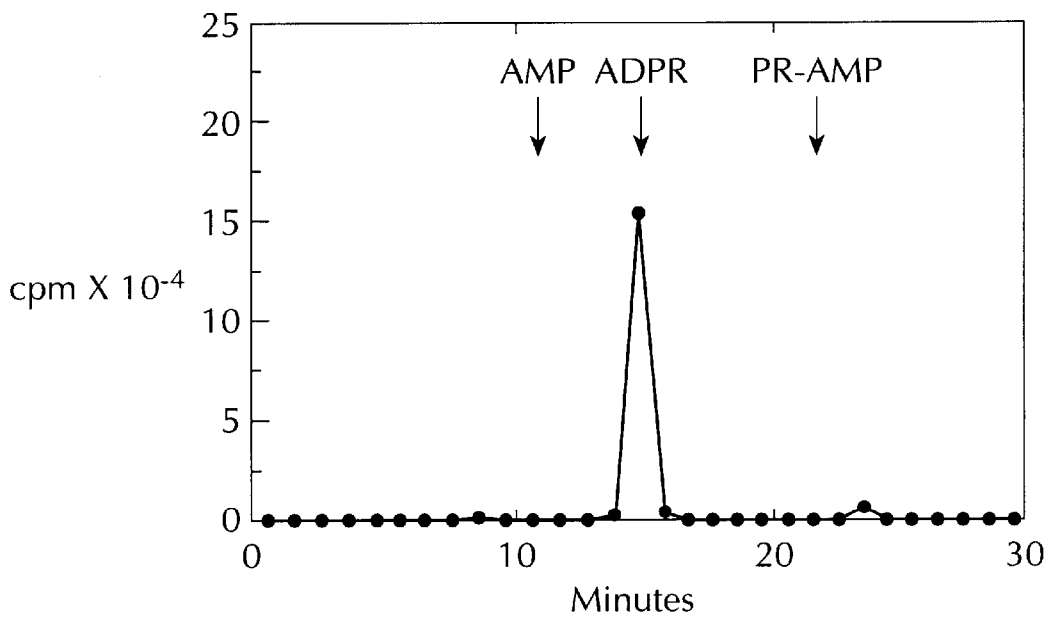


FIG. 7

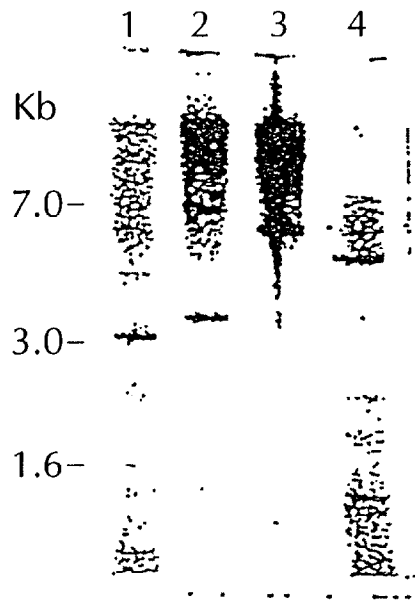


FIG. 8A

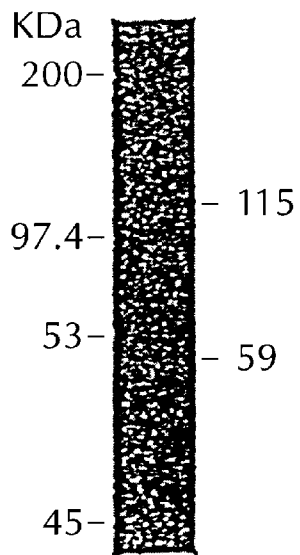


FIG. 8B

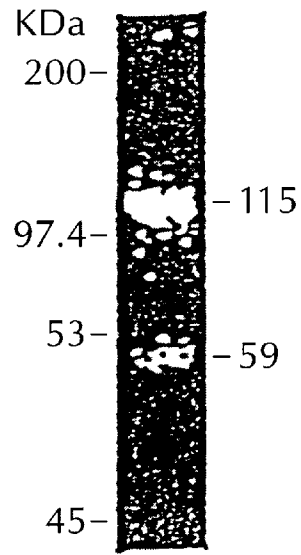


FIG. 10

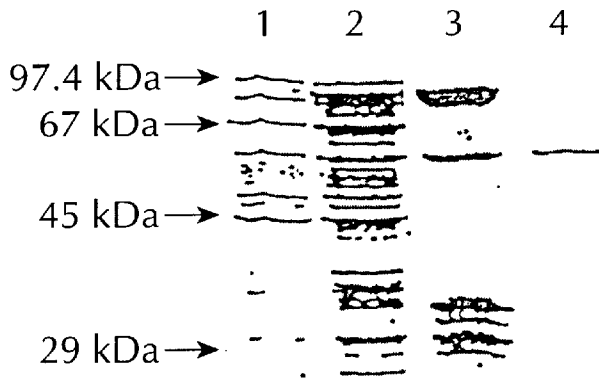


FIG. 12

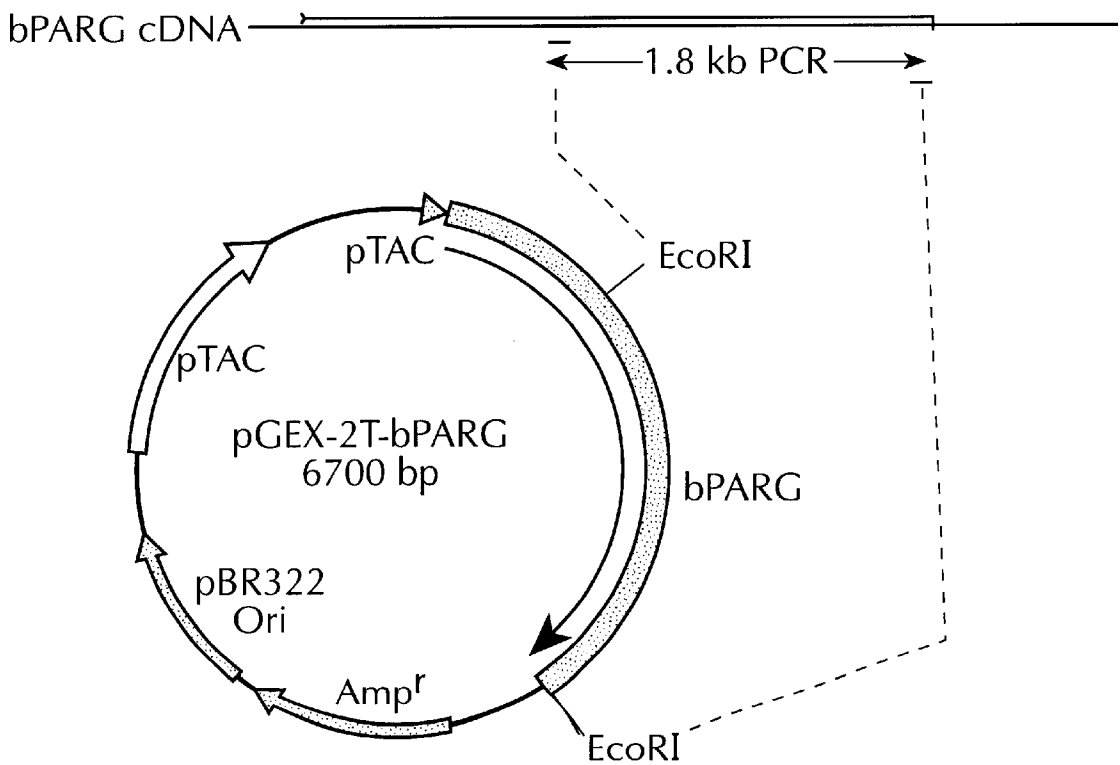


FIG. 11

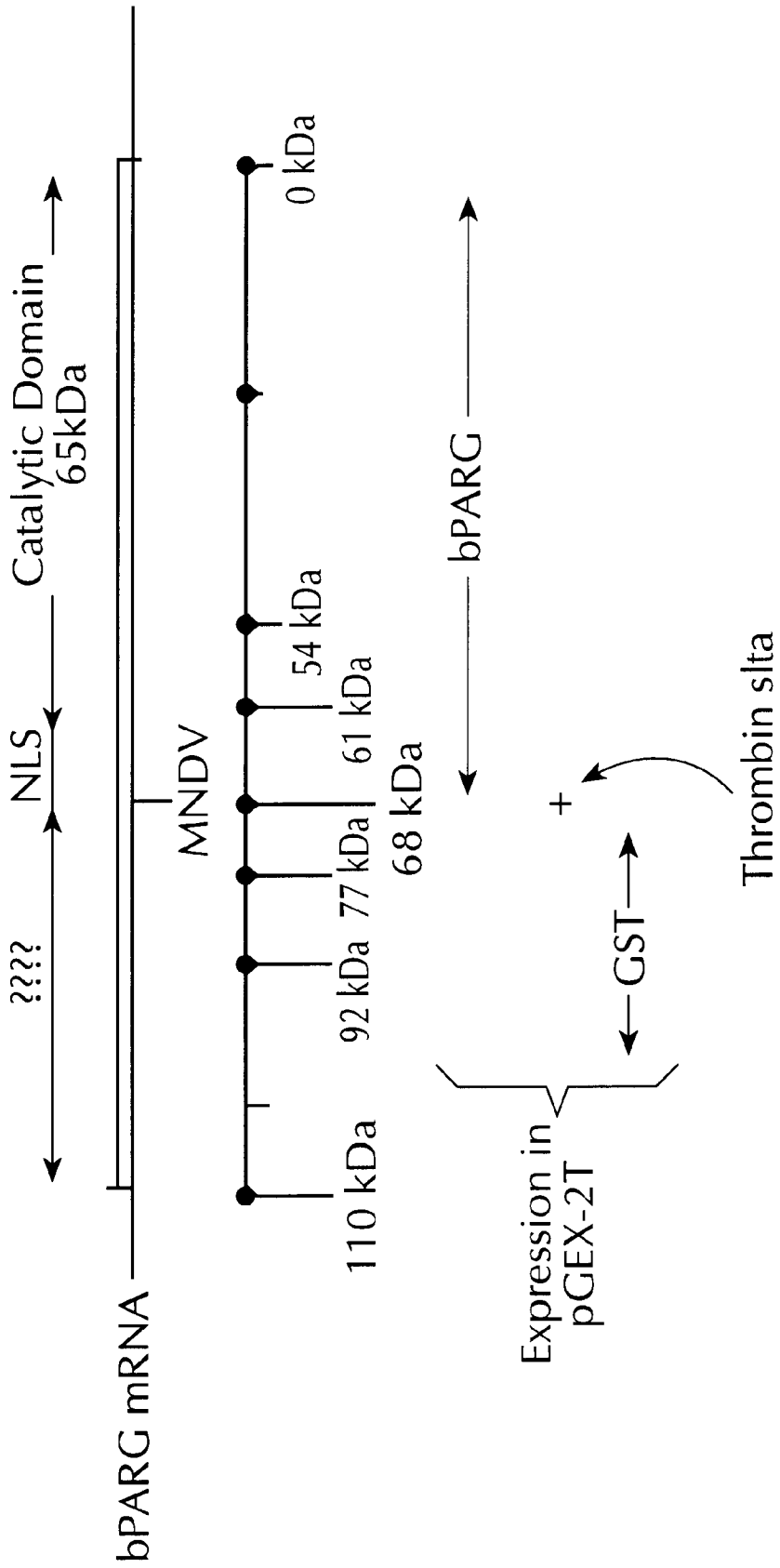


FIG. 13

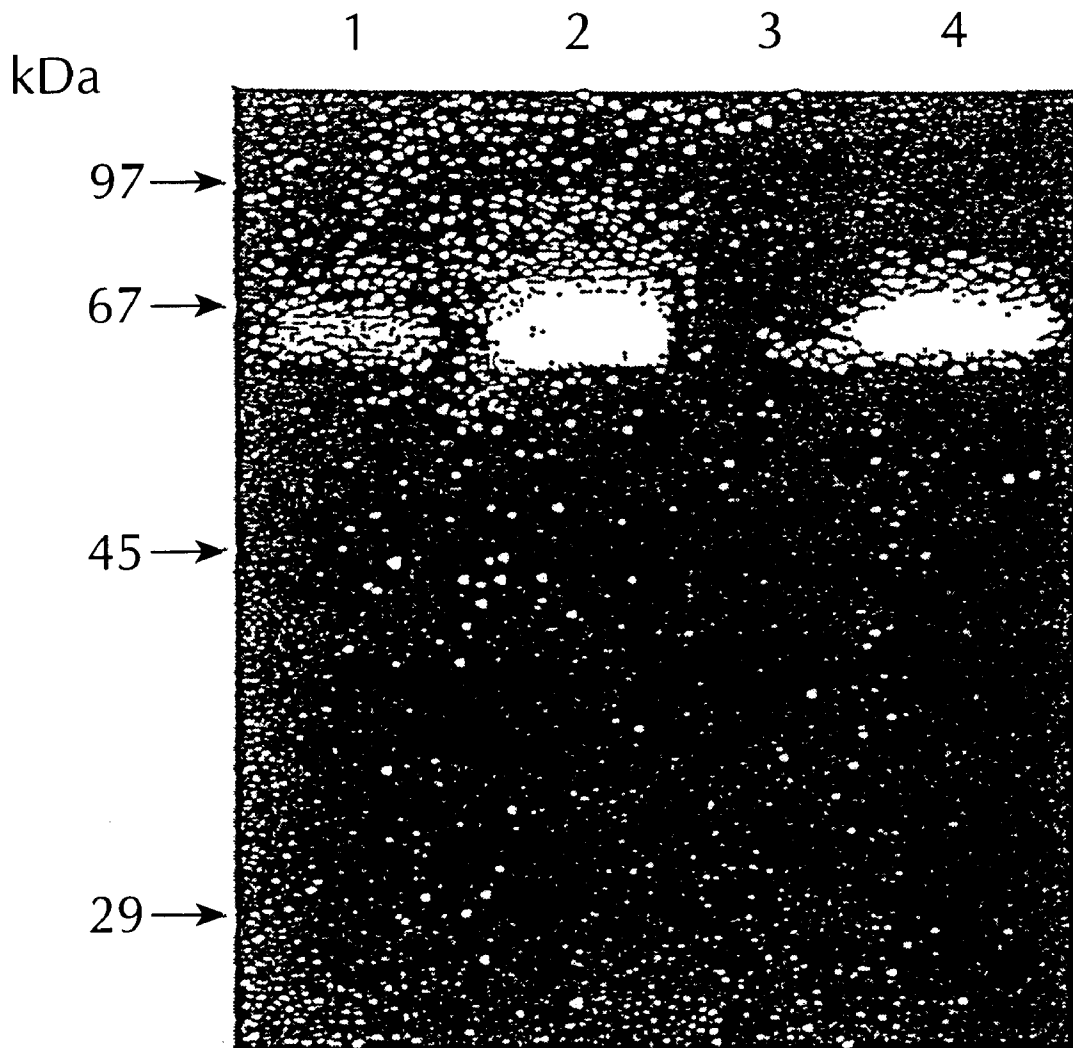


FIG. 14

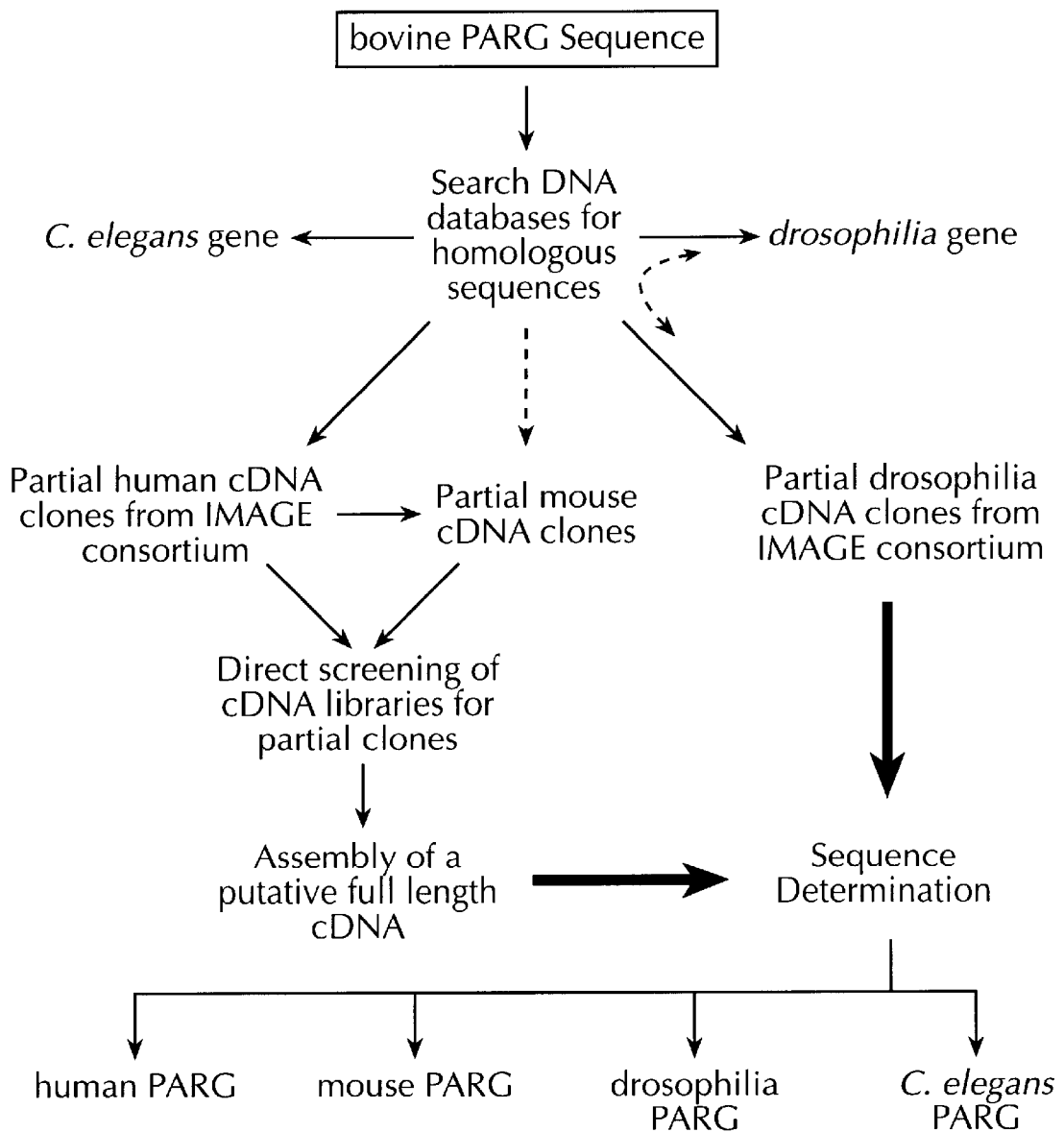


FIG. 15

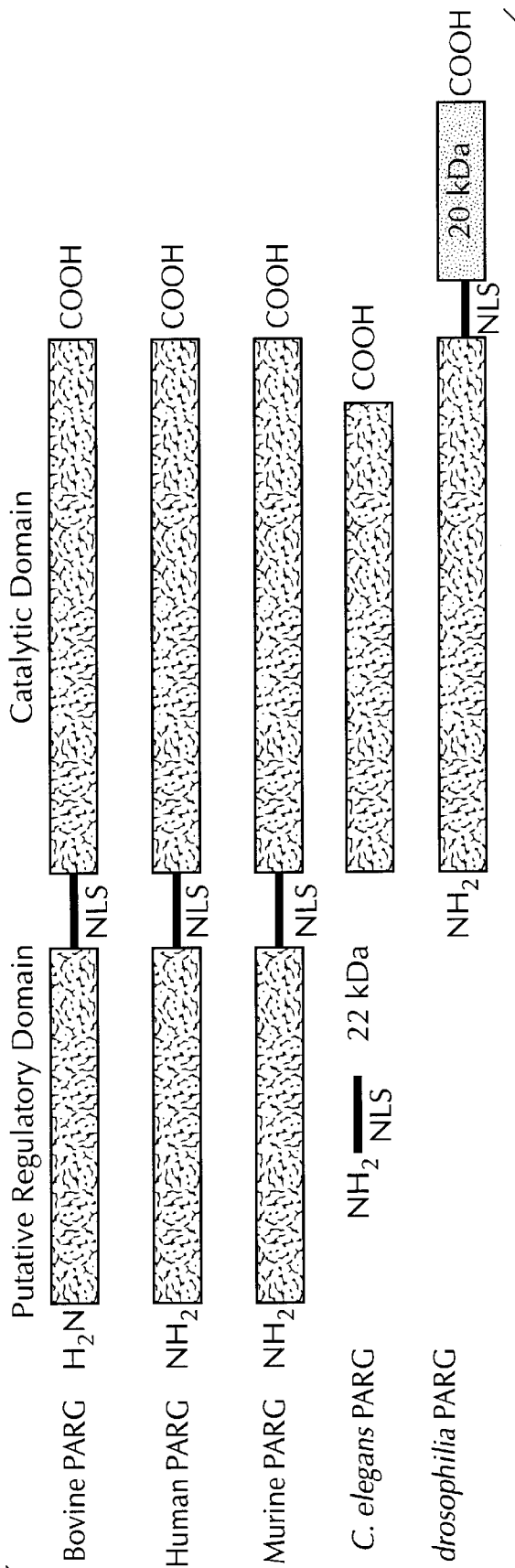
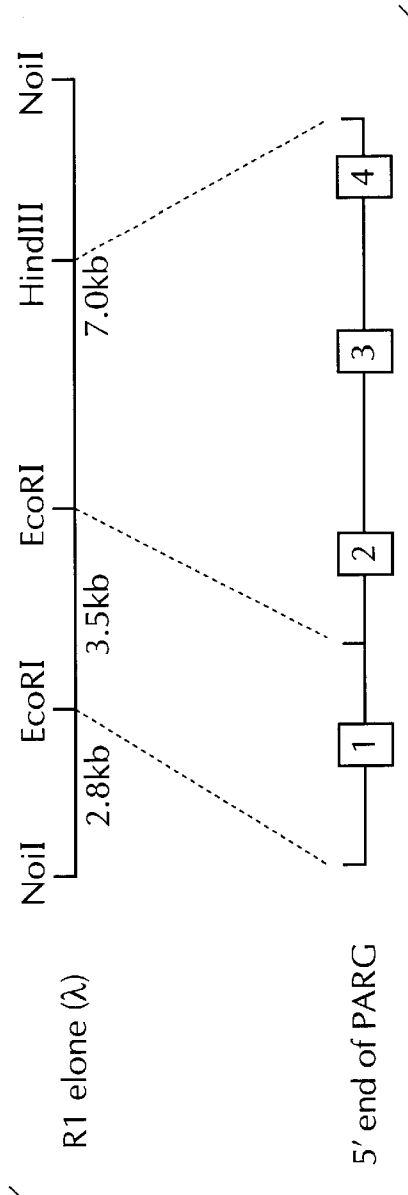


FIG. 20



Bparg 1 MSAGPGCEPCTKRPRWDAAAATSPAAADARSFPGRRVLDKDAVPVQFRVPPSSGCALGRAGQERGSATSL
 Eparg 1 MNAGPGCEPCTKATRWAATTS.PAASDARSFPSRQRRVLDPKDAEVQFRVPPSSPACVPGAGQERGSATSL
 Mparg 1 MSAGPGWEPCTKA.RWGAAGTSAPIASDARSFPGRRVLDPKDAPVQFRVPPSSPACVSGRAGPERGSATHF
 Dparg 1 -----
 Ceparg1 -----

Bparg 74 VFKQKTTITSWMDTKGIKTVEKSLSEXEHHHTREESMMSSVQKDFHYQHMEKLEHVSQGLGFDKSPVEKGTQY
 Eparg 73 VFKQKTTITSWMDTKGIKTVEKSLDSXEHHHTREESMMSSVQKDFHYQHVEKLVHVSQLSLDDKSLTEKSTQY
 Mparg 73 VFKQKTTITSWMDTKGIKTVEK...SXEHHHTREESMMSSVQKDFHYPHKVEKLEHVPQLNLDKSPTEKSSQY
 Dparg 1 -----
 Ceparg1 -----

Bparg 218 LKQHQTAAAMCKWQHGHPSERLLESEPPAVTLVPEQF SHAHVDQSSPKDDHSDTNSSESRDNQQFLTHVKL..
 Eparg 217 LKQHQTAAAMCKWQHGHKHTEQLLSEPPQTVTLVPEQF SHAHIDRSSPQNDHSDTDSEHRDNQQFLTHVKL..
 Mparg 214 LKQQQTASVCKWQHGHKHAEQLLASEPPAGTPLPKQLSHAHIGQSPETDDHSDTDHEEDRDNDQQFLTHVKL..
 Dparg 1 -----
 Ceparg1 ----- MSKKFIELGDPVVTQDEKDYEDYVGVGFAEQVP

Bparg 298 .ANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQEQEETDVVSEPLSDTGSSEPLSDTGLKHAHRLHRQESSL
 Eparg 297 .ANAKQTTEDEHAREAKSHQKCSKCHPGEDCASCQDEIDVVPKSPSLSDVGSSEPLSDTGLKHAHRLHRQESSL
 Mparg 291 .ANTKPTVGDGQ...ARSHCKCSGSRQSKEDCTGCQEQEEVDVLPESPLSDVGAEDIGTGPKHDKLITGQESSL
 Dparg 1 -----
 Ceparg 33 TMKRRKLTEEGNTTESKEDPEEPKSRDVFVSSQSSDESQEDSAENPEIAKEVSENCENLTETLKIENIESLDN

Bparg 396 GHSPPFKESEPEPMDVDNSKHSCQDSEADEETSFGFDE.QEDSSSAQTANKPSRFQPPREADTELKRKRSSAK
 Eparg 395 GHSPPFKESEPEPMDVDNSKHSCQDSEADEETSFGFDE.QEDGSSSQTANKPSRFQARDADIEFRKRYSTK
 Mparg 387 GDSPPFEKESEPEPMDVDNSKHSCQDSEADEETSvFDE.QEDRSS.QTANKLSSCQAREADGDLRKRRLTK
 Dparg 1 MQEFRSHLIFPIFQ.KVIQS.TANRRR.ASASVL.....TNRLGK...ALCLNCARMSKSPDGGISEIE...
 Ceparg106 VTERSEHTLDN...HKSTEPHEE.DVNNKSHIDVAINSDDEDELVLEENHKEMRDGEVQQLS..QDLFADD

FIG. 16A

Bparg 468 GGEIRLHFQFEGGESRAGMN.DVNAKRPGSTSSLNVECRNSKQEGRKDSKITDEFMRVPKAEDKRKEQCEMKH
 Eparg 467 GGEVRLHFQFEGGESRTGMN.DLNAKLPGNISSLNVECRNSKQEGKDSKITDELMRPLPKAEDRRKEQWETKH
 Mparg 354 GGEVRLHFQFE.GEMNAGTS.DLNAKPSGNSSSLNVECRSSKQEGKRDSKITDEFMRISKSEDRRKEQCEVRH
 Dparg 59 .TEEEE.PENLANSL..DDSWRGVSMEAIHRNRQPFLELENLPPVTAAGNLHRVMIQLPIRET..PPR.PIKSPG
 Ceparg 172 QELIEIPGIMKDTTTLQLDITDSEVEETAQKMEMIEETEADSTFVGEDSKATKTVRTSSSSF...LSTVSTCEAP

 Bparg 540 QRTERKIPKYVPPHLSPKDKKWLGTPIEEM..RRMPCRGIRLPPRPSANHTVTIRVDLLRIGEVPK.PFPPTHF
 Eparg 539 QRTERKIPKYVPPPELSPDKKWLGTPIEEM..RRMPCRGIRLPPRPSANHTVTIRVDLLRAGEVPK.PFPPTHY
 Mparg 425 QRTERKIPKYIPPMPLPPEKKWLGTPIEEM..RKMPCRGIHLPSPRPSASHTVTRVDLLRAGEVPK.PFPPTHY
 Dparg 125 X..WDSEEVRLPCAPESKYPRENPDGS.TTIDFRWEMIERALLQPIKTCEELQAAIISIHTTIRDQWEFFRALE
 Ceparg 242 AKGRARMIQKELEKEVIAFTEGNLT....LQPDNLKVDPPDRNIRI.CTIPNFPASQG...KLRED.NRIG..

 Bparg 610 KDLWDNKEVXMPCCSEONLYPVEDENGE.RAAGSRWELIQTALLHRLTRPQNLKDAILKINVAISKKWDFITALI
 Eparg 609 KDLWDNKEVXMPCCSEONLYPVEDENGE.RTAGSRWELIQTALLHKFTRPQNLKDAILKINVAISKKWDFITALI
 Mparg 495 KDLWDNKEVXMPCCSEONLYPVEDENGE.RTAGSRWELIQTALLHKFTRPQNLKDAILKINVAISKKWDFITALV
 Dparg 196 QLLDEELDESETRVFFQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQLASLLANAFCTFPRRNA
 Ceparg 367 PKIVLPQRWRE.FDSRGRR...RDSYFYFKRKLGDGILKCIKTTGYFMFVGLLH.....HMWE..FDPDITIK

 Bparg 682 DFWDXXVLEEAQAHLYSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQLASLLANAFCTFPRRNA
 Eparg 681 DFWDXXVLEEAQAHLYSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQLASLLANAFCTFPRRNA
 Mparg 567 DFWDXXVLEEAQAHLYSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQLASLLANAFCTFPRRNA
 Dparg 269 .KMKSEISSYPDIHFNRLLFEGRSSRKPPELKLTLFCIFRRV..TERDASNVTGVVTVFRRSGLPEE.L.IDWS
 Ceparg 424 ...LPAL.EMYYKEMSELVGREEVLEKFAFARVARIAKTAEDILPERIYRLVGDVE.SATLSHKQCAALVARMF

 Bparg 755 .KMKSEISSYPDIHFNRLLFEGRSSRKPPELKLTLFCIFRRV..TEKK...PTGLVTFIRQS.L.ED.F.PEWE
 Eparg 754 .KMKSEISSYPDIHFNRLLFEGRSSRKPPELKLTLFCIFRRV..TEKK...PTGLVTFIRQS.L.ED.F.PEWE
 Mparg 387 .KMKSEISSYPDIHFNRLLFEGRSSRKPPELKLTLFCIFRRV..TEKK...PTGLVTFIRQS.L.ED.F.PEWE
 Dparg 337 QSAAPLGDVPLEVDAEGTIEDEGIGLQVDFAHKYLGGVLGHCVCQVEEIRFVICPELLVGLKLFTECLRPFEA
 Ceparg 491 FA.....RPDSPFS.....FCRILSSDKSICVEKLLKFLFTIF.....DKMSMDPPDGAVSF.RLTKMD

FIG. 16B

Bparg 817 RCEKLL..TRLEVIYEGTIEGNQGMLQVDFAHRFVGGVTSAGLVQEEIRFLIHPELIVSRLFIEVLDHNEC
 Eparg 816 RCEKPL..TRLEVIYEGTIEENGQMLQVDFAHRFVGGVTSAGLVQEEIRFLIHPELII SRLFIEVLDHNEC
 Mparg 702 RCEKPL..TRLEVIYEGTIEGNRGRMLQVDFAHRFVGGVTSAGLVQEEIRFLIHPELIVSRLFIEVLDHNEC
 Dparg 410 LVMLGAERYSNYGTIAGSFEWS....GNFEDSTP.RDSSGRRQTAIVAIDALHFA.QSHH....QYREDL
 Ceparg 543 KDIFNEEW.XDXXLLRSLPEVEFFDEMLIEDTAL.CTQVDFAHHEHLGGVNLHGVSQVEEIRFLMCPMMVGMLL

 Bparg 888 LIITGTEQISEYTGIAETYRWA....RSHEDRSE.RDDWQRRTTEIVAIDALHFR.RYLD....QFVPEKIR
 Eparg 887 LIITGTEQISEYTGIAETYRWS....RSHEDRSE.RDDCERRCTEIVAIDALHFR.RYLD....QFVPEKMR
 Mparg 773 LIITGTEQISEYTGIAETYRWA....RSHEDRSE.KDDWQRRCTEIVAIDALHFR.RYLD....QFVPEKVR
 Dparg 472 MERELHKAYIGFVHMMVTPP..PGVATGNWGCAGFGGDSYLLKALLQLMVCACLGRPLAYYTFGNVEFRDDF
 Ceparg 614 CEKMKQLEAISIVGAYVFSSTYGYGHTLKWAELOPNHSRQNTNEFRDRFRGRLRVETIAIDAILFKGSKLDCQT

 Bparg 951 RELHKAYCGFLRPGVSSSEHLSAVATGNWGCAGFGGDARLKALIQYLAAAVAERDVVFTFGDSELMRDIYSMH
 Eparg 950 RELHKAYCGFLRPGVSSSEHLSAVATGNWGCAGFGGDARLKALIQYLAAAVAERDVVFTFGDSELMRDIYSMH
 Mparg 835 RELHKAYCGFLRPGVSPSEHLSAVATGNWGCAGFGGDARLKALIQYLAAAVAERDVVFTFGDSELMRDIYSMH
 Dparg 543 HEMWLLFRNDGTTVQQ.LMS.IILRSYRLIKEKSSKEPRENKASKKLYDFI..KEELKKVRDVPGEAS
 Ceparg 687 EQLNKANIIIREMKKASIGFMSQGPKFTHIP.IVTGWGCGAFNGDKPLKFI IQVIAAGVADRPLHFCSFGEPE

 Bparg 1021 TFLTERKLTVGE.VYKLLRLRYNNEECRCNCSTPGP.....DIKLYPFYIHAVESCTQTNNQPGQRTGA-----
 Eparg 1020 IFLTERKLTVGD.VYKLLRLRYNNEECRCNCSTPGP.....DIKLYPFYIHAVESCAETADHSGQRTGT-----
 Mparg 907 TFLTERKLDVGK.VYKLLRLRYNNEECRCNCSTPGP.....DIKLYPFYIHAVESSAETDMPGQKAGT-----
 Dparg 611 AEAGSSRVAGLGEKSETSAKSSPELNKQPARPQITITQQSTDLLPAQLSQDNNSSEDDQALLMLSDDEEA
 Ceparg 759 LAAXCKKIIERMKQKDVTLGKSCFSIFS-----

 Bparg 1087 -----
 Eparg 1086 -----
 Mparg 969 -----
 Dparg 684 NAMEAASLEAKSSVEISNSSSTTSKTSSTATKSMGSGRQLSLEMLDTHYEKGSASKRPRKSPNCskaEG
 Ceparg 787 -----

FIG. 16C

Bparg 1	-----
Eparg 1	-----
Mparg 1	-----
Dparg 410	TDKDEKDDIVD
Ceparg 543	-----
Bparg 101	-----
Eparg 100	-----
Mparg 97	-----
Dparg 472	-----
Ceparg 614	-----
Bparg 201	-----
Eparg 200	-----
Mparg 197	-----
Dparg 543	-----
Ceparg 687	-----

FIG. 16D

FIG. 17

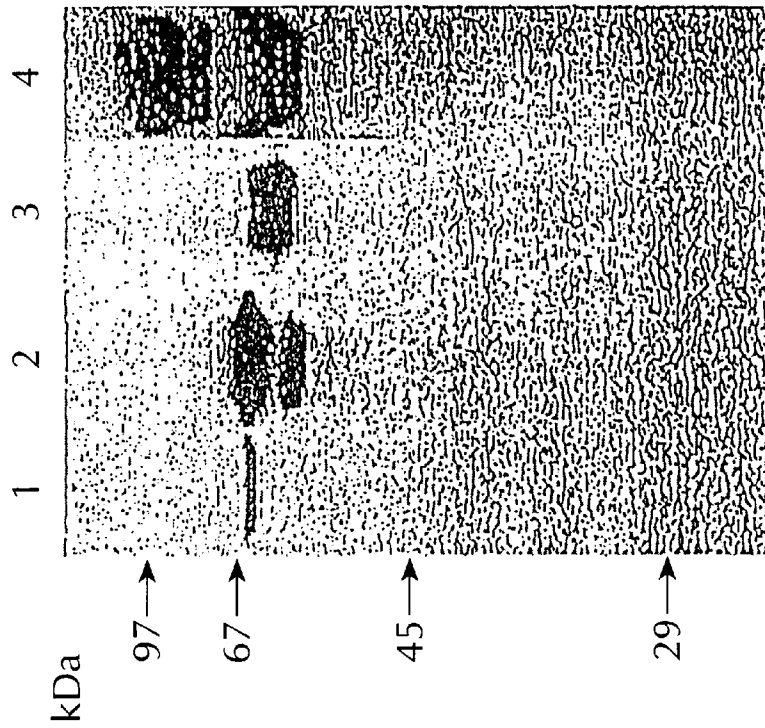


FIG. 18A

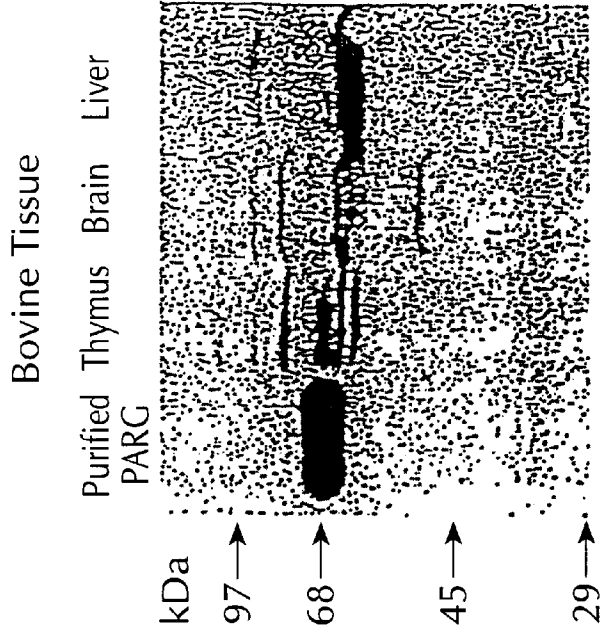


FIG. 19

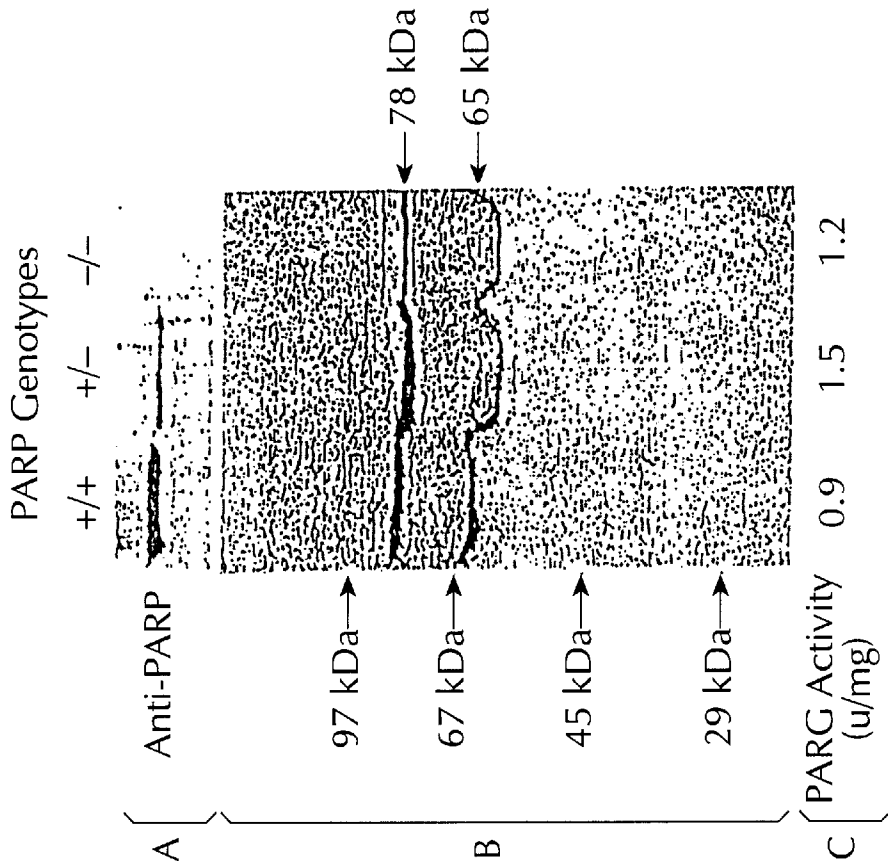


FIG. 18B

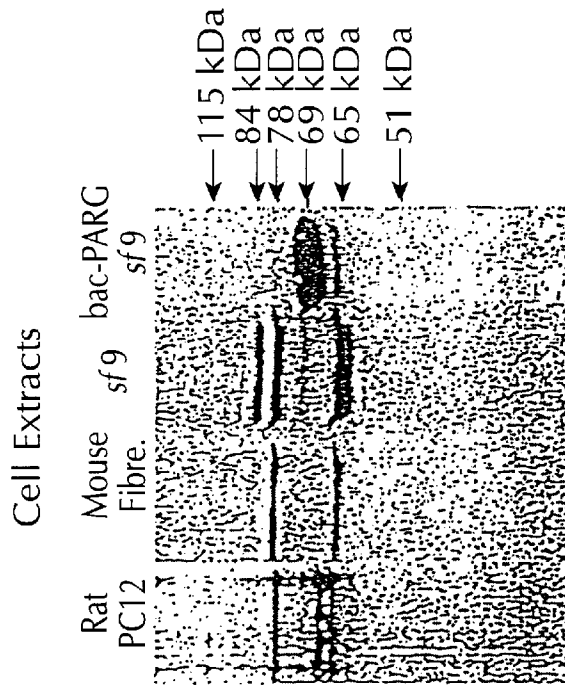
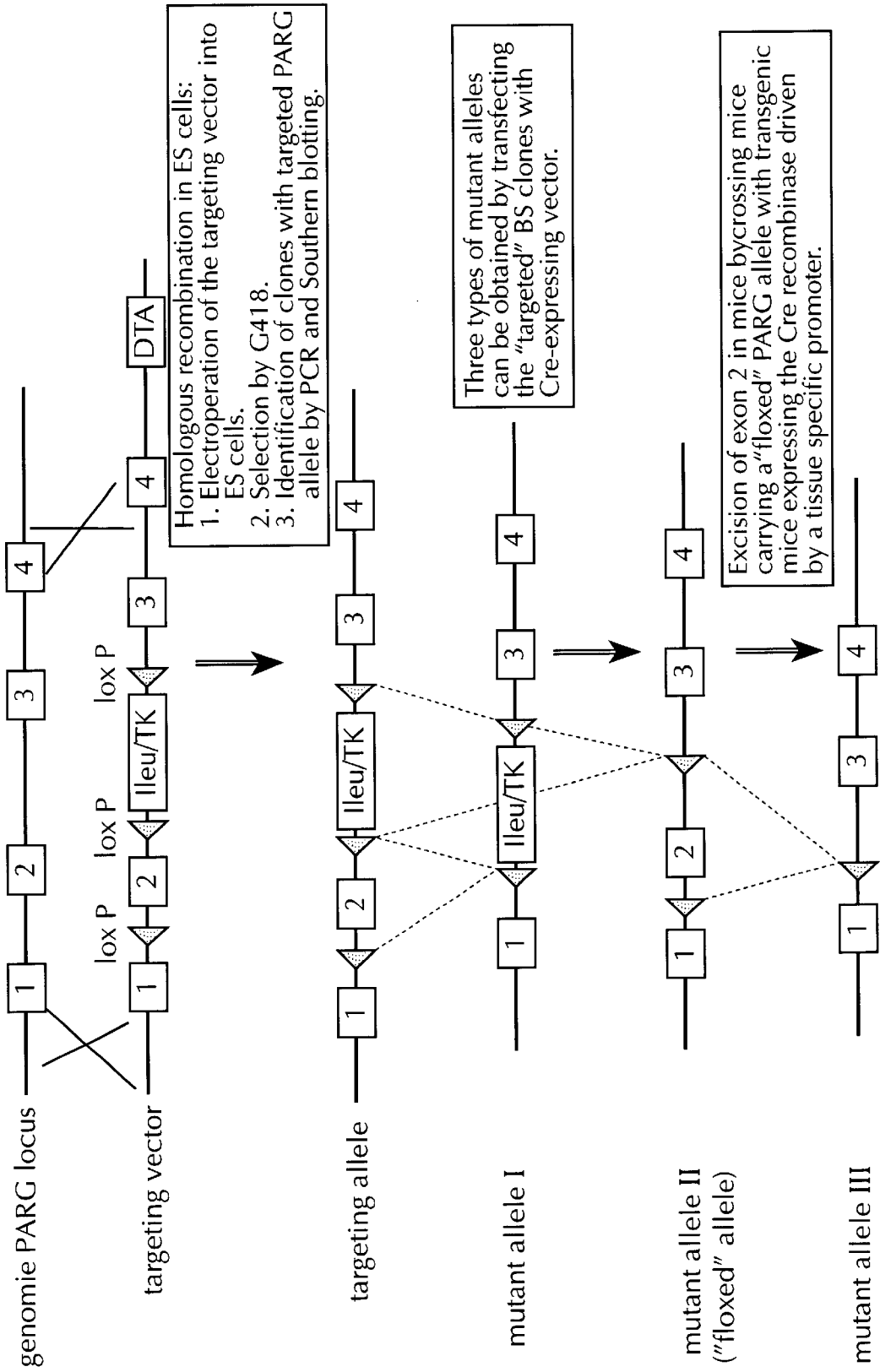


FIG. 21



GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZYMES, THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE THEREWITH

CROSS REFERENCE TO RELATED APPLICATION

This application is a divisional of U.S. application Ser. No. 09/302,812 filed Apr. 30, 1999, which claims benefits of U.S. provisional Application No. 60/083,768 filed May 1, 1998. The entire disclosure of U.S. application Ser. No. 09/302,812 and U.S. Provisional Application No. 60/083,768 is incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

The present invention was supported in part by the National Institutes of Health (Grant CA43894). The United States Government may have certain rights in the invention.

TECHNICAL FIELD

The present invention relates to poly(ADP-ribose) glycohydrolases (PARGs) and peptides having poly(ADP-ribose) glycohydrolase activity. In addition, the invention also relates to antibodies, including monoclonal antibodies and antibody fragments, that have specific interaction with epitopes present on poly(ADP-ribose) glycohydrolases. Methods of treatment and diagnosis using the poly(ADP-ribose) glycohydrolases, and antibodies specific for poly(ADP-ribose) glycohydrolases are disclosed. The present invention has implications for the treatment of neoplastic disorder, reperfusion following ischemia, neurological disorders, and related conditions.

BACKGROUND OF THE INVENTION

Genomic damage, if left unrepaired, can lead to malignant transformation, or cell death by senescence (aging), necrosis or apoptosis. Among the variables that can affect the ultimate biological consequence of DNA damage to a particular cell are (i) the amount, type, and location of the DNA damage and (ii) the efficiency and bioavailability of the cellular DNA repair mechanism.

The activation of poly(ADP-ribose) polymerase (PARP) by DNA strand breaks is often one of the first cellular responses to DNA damage. PARP catalyzes the conversion of nicotinamide adenine dinucleotide (NAD) to multi-branched polymers containing up to 200 ADP-ribose residues. Increases in polymer levels of more than 100-fold may occur within minutes of DNA damage. Once synthesized, polymers are rapidly turned over, being converted to free ADP-ribose by the action of poly(ADP-ribose) glycohydrolase (PARG) (1). An ADP-ribosyl protein lyase has been proposed to catalyze removal of protein-proximal ADP-ribose monomers (2). FIG. 1 illustrates these processes schematically.

The process of activating PARP upon DNA damage can rapidly lead to energy depletion because each ADP-ribose unit transferred by PARP consumes one molecule of NAD, which in turn, requires six molecules of ATP to regenerate NAD. Additionally, NAD is a key carrier of electrons needed to generate ATP via electron transport and oxidative phosphorylation or by glycolysis. The overactivation of PARP due to substantial DNA damage can significantly deplete the cellular pools of NAD and ATP (3). ADP-ribose polymer

metabolism, and thus PARP and PARG have been linked to the enhancement of DNA repair (4), limitation of malignant transformation (5), enhancement of necrotic cell death (6), and involvement in programmed cell death (7). To date, studies of the structure and function of the enzymes of ADP-ribose polymer metabolism have been mainly limited to PARP (8). Little is known about the function and regulation of PARG.

BRIEF SUMMARY OF THE INVENTION

As embodied and broadly described herein, the present invention is directed to nucleic acids molecules, peptides, methods, vectors and antibodies that are related to the poly(ADP-ribose) glycohydrolase (PARG) enzyme.

One embodiment of the invention is directed to an isolated and purified nucleic acid molecule or nucleic acid molecule analog comprising a sequence that encodes a polypeptide having poly(ADP-ribose) glycohydrolase (PARG) activity. The nucleic acid molecule may encode the complete full-length PARG gene or a fragment of the PARG gene. The nucleic acid molecule may be DNA, RNA or peptide nucleic acid (PNA). The nucleic acid molecule can be linear, such as, for example, an isolated fragment or a linear phage DNA. In addition, the isolated nucleic acid molecule may be circular, such as for example in a plasmid. The nucleic acid molecule may also be a single stranded DNA or RNA such as the single stranded DNA or RNA in a single stranded DNA virus or single stranded RNA virus. The nucleic acid molecule may be of yeast, insect or mammalian origin.

The nucleic acid molecule of the invention, may be of mammalian origin, such as, for example of bovine or murine origin. In a preferred embodiment of the invention, the nucleic acid molecule may be of human origin. While the sequence of the nucleic acid molecule is of mammalian origin, the nucleic acid molecule may be replicated in another organism such as an insert in a viral genome, a plasmid in a bacterium or a 2-micron plasmid in a yeast.

Preferably, the nucleic acid molecule has, a high degree of sequence similarity with a sequence shown in SEQ ID NO: 1 (Genbank Accession Number U78975), SEQ ID NO: 3 (Genbank Accession Number AF005043), SEQ ID NO: 5 (Genbank Accession Number AF079557), SEQ ID NO: 7 (Genbank Accession Number AF079556) or SEQ ID NO: 9 (Genbank Accession Number CEF20C5). The high degree of sequence similarity may be, for example, about 70%, preferably about 80%, even more preferably about 90% and most preferably substantially identical such as for example about 100% identity.

The nucleic acid molecule that encodes a polypeptide having poly(ADP-ribose) glycohydrolase (PARG) activity may be single or double stranded nucleic acid molecule of any length such as, for example, about 20 bases in length, about 30 bases in length, about 40 bases in length, about 50 bases in length, about 100 bases in length, about 200 bases in length, about 500 bases in length, about 1000 bases in length, about 1500 bases in length, about 2000 bases in length, about 3000 bases in length. It is understood that "bases" in this patent application means "basepairs" when referring to double stranded nucleic acid molecules and bases when referring to single stranded nucleic acid molecules. In a preferred embodiment of the invention, the nucleic acid molecule may be at least about 1000 base or basepairs long and have at least about 80% sequence similarity with a sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9.

In one embodiment of the invention, the nucleic acid molecule may have sequence similarity to one region of the PARG sequence. The region may be, for example, from about base residue 2113 to about residue 3105 of SEQ ID NO: 3. Alternatively, the region may be, from residue 1240 to about residue 3105 of SEQ ID NO: 3 or from residue 175 to about residue 3105 of SEQ ID NO: 3.

Another embodiment of the invention is directed to the expression and overexpression of PARG in a cell. Expression vectors may mediate the expression of a polypeptide with poly (ADP-ribose) glycohydrolase (PARG) enzyme activity. Expression systems and expression vectors are known in the art. For example, one expression vector may comprise a regulatory sequence which is operatively linked to a nucleotide sequence at least about 1000 base pairs in length, which has at least 70% sequence similarity with a sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9. In a preferred embodiment, the sequence similarity is at least about 80% identity, more preferably at least about 90% identity and most preferably about 100% identity. The expression vector may be any expression vector that is capable of directing expression of a gene in a host cell including, prokaryotic, eukaryotic, mammalian and viral vector. Examples of such vectors include pCMV-Script cytomegalovirus expression vectors for expression in mammalian cells, pESP and pESC vectors for expression in *S. pombe* and *S. cerevisiae*, pET vectors for expression in bacteria, pSPUTK vectors for high-level transient expression, and pBac and pMbac vectors for expression in fall army worm (SF9) cells. Such vectors are available commercially from suppliers such as, for example, Invitrogen (Carlsbad, Calif.) or Stratagene (La Jolla, Calif.). In the use of viral vectors, it is understood that defective viral vectors—vectors that are genetically engineered to deliver a gene or gene product to a host but which cannot replicate in a host is preferred. Procedures for the practice of in vitro and in vivo expression are well known to those of skill in the art and are further available with the specific expression products and cell lines from commercial suppliers.

Another embodiment of the invention is directed to a host cell transformed with a vector containing a nucleic acid molecule with a sequence that encodes a polypeptide having poly(ADP-ribose) glycohydrolase (PARG) activity. The host cell may be any eukaryotic or prokaryotic cell such as, for example a human, murine, rattus, bovine, insect, yeast or bacteria. Specific cell lines are well known to those of skill in the art and are available from suppliers such as the American Tissue Type Collection (ATCC, Manassas, Va.) and Stratagene (La Jolla, Calif.) and the like. A preferred embodiment of the invention is directed to cells transformed with the PARG expression vector which shows an elevated level of PARG relative to non-transformed cells. Especially preferred are cells transformed with an inducible PARG expression vector that have normal or slightly elevated PARG levels before induction and have significantly elevated PARG levels after induction.

An embodiment of the invention is directed to an isolated protein having poly(ADP-ribose) glycohydrolase (PARG) activity. The protein may comprise an amino acid sequence with at least 70% sequence similarity with a sequence shown in SEQ ID NO: 2 (Genbank Accession Number U78975), SEQ ID NO: 4 (Genbank Accession Number AF005043), SEQ ID NO: 6 (Genbank Accession Number AF079557), SEQ ID NO: 8 (Genbank Accession Number AF079556), or SEQ ID NO: 10 (Genbank Accession Number CEF20C5). The sequence similarity is preferably at least about 80%,

more preferably at least about 90% and most preferably substantially identical with a sequence shown in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In a preferred embodiment of the invention, the preferred isolated protein having poly(ADP-ribose) glycohydrolase (PARG) activity and has a molecular weight greater than about 100 kDa.

Another embodiment of the invention is directed to an oligonucleotide which is greater than about 10 bases in length and less than about 1000 bases in length which is complementary to a sequence shown SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. The oligonucleotide may be, for example, greater than about 20 bases in length, greater than about 30 bases in length, greater than about 40 bases in length, greater than about 50 bases in length, greater than about 100 bases in length, greater than about 200 bases in length or greater than about 300 bases in length. The oligonucleotide, which may be optionally labeled with a detectable marker, may be DNA, RNA or PNA. A detectable marker may be, for example, a radioactive isotope such as ³²P or ¹²⁵I, an epitope such as FLAG.

One preferred oligonucleotide is an antisense oligonucleotide directed to the mRNA of PARG. Antisense oligonucleotide as a method of suppression is well known to those in the art. For example, the phosphorothioate oligonucleotide, ISIS 2922, has been shown to be effective against cytomegalovirus retinitis in AIDS patients (9). It is thus well known that oligonucleotides, when administered to animals and humans, can have a useful therapeutic effect. In a preferred embodiment, the oligonucleotide is at least about 10 nucleotides in length, such as, greater than about 20 bases in length, greater than about 30 bases in length, greater than about 40 bases in length, greater than about 50 bases in length, greater than about 100 bases in length, greater than about 200 bases in length or greater than about 300 bases in length. In another preferred embodiment, the oligonucleotide has a ribozyme activity.

Another embodiment of the invention is directed to an isolated polypeptide of at least 6 amino acid residues in length and having a molecular weight less than about 65 kDa, which has at least about 80% sequence similarity with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. The polypeptide may be, for example, at least about 10 amino acids in length, at least about 20 amino acids in length, at least about 30 amino acids in length, at least about 40 amino acids in length, at least about 50 amino acids in length, at least about 75 amino acids in length, at least about 100 amino acids in length, at least about 150 amino acids in length, at least about 250 amino acids in length or at least about 500 amino acids in length or more.

In a preferred embodiment, the polypeptide has a molecular weight less than about 40 kDa and has at least about 90% sequence similarity with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. The polypeptide preferably has poly(ADP-ribose) glycohydrolase (PARG) activity or is immunogenic and elicits antibodies immunoreactive with a poly(ADP-ribose) glycohydrolase (PARG) enzyme. In a more preferred embodiment, the polypeptide comprises an amino acid sequence substantially identical with SEQ ID NO: 4 from about residue 647 to about residue 977.

Another embodiment of the invention is directed to an isolated polypeptide of at least 10 amino acid residues in length and which has at least about 80% sequence similarity

with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. Preferably, the polypeptide is at least about 20 amino acids in length, such as, for example at least about 30 amino acids, about 40 amino acids, about 50 amino acids, about 100 amino acids, about 200 amino acids and about 500 amino acids in length.

Another embodiment of the invention is directed to an antibody immunoreactive with an isolated polypeptide of at least about 6 amino acid residues in length and having a molecular weight less than about 65 kDa, which has at least about 80% sequence similarity with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. In a preferred embodiment, antibody is immunoreactive with a polypeptide with a molecular weight less than about 40 kDa and has at least about 90% sequence similarity with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. In another preferred embodiment, the antibody is immunoreactive with a polypeptide comprising an amino acid sequence substantially identical with SEQ ID NO: 4 from about residue 647 to about residue 977.

Another embodiment of the invention is directed to a method of detecting a polypeptide having PARG activity comprising the steps of contacting the polypeptide with an antibody immunoreactive with an isolated polypeptide of at least about 6 amino acid residues in length and having a molecular weight less than about 65 kDa, which has at least about 80% sequence similarity with a sequence shown in any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10, and determining whether the antibody immunoreacts with the polypeptide.

Another embodiment of the invention is directed to a method of preventing, treating, or ameliorating a disease condition or disorder in an individual comprising the step of administering a therapeutically effective amount of a poly(ADP-ribose) glycohydrolase (PARG) inhibitor or activator to the individual. The disease condition or disorder may be any condition associated with responses to DNA damage, examples of which include a neoplastic disorder, a myocardial infarction, a vascular stroke or a neurodegenerative disorder. The PARG inhibitor or activator may be a small molecule inhibitor or activator of PARG or may be an antisense oligonucleotide that can hybridize in vivo to messenger RNA encoded by a PARG gene. PARG based treatment may be directed to new methods for preventing, treating or ameliorating disorders associated with DNA damage. These disorders include neoplastic disorders, inborn genetic errors, myocardial infarctions, vascular strokes, aging, and neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, Parkinson's disease, and neurotoxicity generally.

Another embodiment of the invention is directed to the identification of novel PARG modulators which can activate or inhibit DNA repair and/or apoptosis. A PARG modulator is a compound that can activate or inhibit PARG. These modulators are preferably more efficacious and do not have the known side effects of present modulators. One method of identifying an agent that inhibits or activates poly(ADP-ribose) glycohydrolase (PARG) activity comprise the steps of providing a liquid medium that contains a polypeptide having PARG activity contacting the polypeptide with a candidate agent, in the presence of a reference compound having affinity for the polypeptide, under predetermined assay conditions, and determining the affinity of the candidate agent for the polypeptide relative to the reference

compound. Thus, the modulation activity of the candidate agent relative to the reference compound is determined. In this method, the polypeptide may be immobilized on a solid support. Further, the polypeptide may be generated in vitro by culturing a cell transformed with a nucleic acid molecule encoding PARG under conditions effective to express the polypeptide.

Another embodiment of the invention is directed to a method of identifying a mutant PARG allele in an individual comprising the step of obtaining genomic material from the individual; digesting the genomic material with a restriction enzyme having a recognition site inclusive of the mutant allele; fractionating the restriction fragments obtained from the digestion; and comparing the fractionation pattern with that obtained for a normal allele, thereby determining the presence or absence of the mutant allele. The fractionation step may be performed with electrophoresis.

Another embodiment of the invention is directed to a method of identifying a mutant PARG allele in an individual comprising the steps of hybridizing an oligonucleotide with genomic material from the individual, which oligonucleotide hybridizes under predetermined hybridization conditions to a region immediately 5' of a predetermined mutation site in the PARG alleles with the 3' terminus of the oligonucleotide complementary to an unmutated PARG allele; extending the oligonucleotide using PCR amplification; and determining the degree to which extension occurs, thereby determining the presence or absence of the mutant allele. The PCR extension reaction may be performed at a temperature above about 50° C. The determination may be performed by conducting electrophoresis (using for example, acrylamide at about 4% to about 10% or agarose and low melting temperature agarose from about 0.8% to about 4%) on the products of PCR amplification.

Another embodiment of the invention is directed to a method of screening molecules for PARG modulating activity (inhibition or activation) comprising the steps of providing a purified PARG enzyme; assaying the enzyme in the presence of a molecule to be screened; and comparing the activity of the PARG enzyme in the presence of the molecule to the activity of the PARG enzyme in the absence of the molecule.

Another embodiment of the invention is directed to a method of gene therapy comprising the step of delivering an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme to a cell to be treated. In the method, the oligonucleotide may have a sequence complementary to a sequence encoding a C-terminal portion of a PARG enzyme. Further, in the gene therapy method, the oligonucleotide may further comprise a ribozyme.

Another embodiment of the invention is directed to a method of delivering to a cell surface, an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme to a cell to be treated. In the method, the oligonucleotide may have a sequence complementary to a sequence encoding a C-terminal portion of a PARG enzyme. Further, in the method, the oligonucleotide may further comprise a ribozyme. The portion of a polynucleotide encoding a PARG enzyme may be, for example, the polynucleotide encoding the N terminus third of PARG, the middle third of PARG, or the C terminus third of PARG. The portion of a polynucleotide may encode a smaller part of PARG such as the N terminus 10% of PARG, the C terminus 10% of PARG, or any 10% portion in between such as from 10% to 20%, from

20% to 30%, from 30% to 40%, from 40% to 50%, from 50% to 60%, from 60% to 70%, from 70% to 80%, from 80% to 90%. The percent value used means a percent of the linear amino acid sequence. Thus, for a 1000 amino acid protein, the N terminus 10 percent is from amino acid 1 to 100; 10% to 20% percent would be from amino acid 100 to 200 and so on. For a 970 amino acid protein, the N terminal 10% would be from amino acid 1 to 97; 10% to 20% would be from amino acids 98 to 194 amino acids.

Another embodiment of the invention is directed to a method of sensitizing a cell to a chemotherapeutic agent comprising the step of contacting the cell with a molecule that modulates the activity of a PARG enzyme. The molecule may be an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme. For example, the oligonucleotide may have a sequence complementary to a sequence encoding a C-terminal portion of a PARG enzyme. The portion of a polynucleotide encoding a PARG enzyme may be, for example, the polynucleotide encoding the N terminus third of PARG, the middle third of PARG, or the C terminus third of PARG. The portion of a polynucleotide may encode a smaller part of PARG such as the N terminus 10% of PARG, the C terminus 10% of PARG, or any 10% portion in between such as from 10% to 20%, from 20% to 30%, from 30% to 40%, from 40% to 50%, from 50% to 60%, from 60% to 70%, from 70% to 80%, from 80% to 90%. The oligonucleotide may further comprise a ribozyme. The method may be used, for example, as a method of treating a diseased cell characterized by the presence of DNA strand breaks. In the treatment, the cell is contacted with a molecule that modulates an enzymatic activity of a PARG enzyme.

Another embodiment of the invention is directed to a pharmaceutical composition comprising an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme. The produced molecule may be an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme. For example, the oligonucleotide may have a sequence complementary to a sequence encoding a C-terminal portion of a PARG enzyme. The oligonucleotide may comprise a ribozyme activity.

Another embodiment of the invention is directed to a virus that causes the production of an oligonucleotide having a sequence complementary to a polynucleotide encoding a PARG enzyme. This may be, for example, a viral vector which after the infection of a host cell, causes the production of an antisense RNA of PARG. The molecule may be an oligonucleotide having a sequence complementary to at least a portion of a polynucleotide encoding a PARG enzyme. For example, the oligonucleotide may have a sequence complementary to a sequence encoding a C-terminal portion of a PARG enzyme. The oligonucleotide may further comprise a ribozyme activity.

Other embodiments and advantages of the invention are set forth, in part, in the description that follows and, in part, will be obvious from this description and may be learned from the practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts the cellular biochemical process that occurs after DNA damage.

FIG. 2 depicts the SDS-PAGE analysis of purified bovine thymus PARG.

FIG. 3 depicts the alignment of the DNA sequences of two PCR products and eight λgt11 cDNA clones used to identify the cDNA coding for bovine PARG.

FIG. 4 depicts a northern blot analysis of bovine kidney cells mRNA transcripts.

FIG. 5 depicts an alignment of the putative bipartite NLS of bovine, human, and murine PARG and comparison with the bipartite NLS of PARP from different organisms.

FIG. 6 depicts expression of bPARG enzyme activity in *E. coli* (10).

FIG. 7 depicts a Southern blot analysis of bovine DNA probed with PARG cDNA.

FIG. 8 depicts activity gel autoradiogram of *E. coli* expressed bovine PARG.

FIG. 9 depicts the analysis by anion exchange HPLC of material released from ADP-ribose polymers by PARG action.

FIG. 10 depicts the SDS-PAGE analysis of the purification of *E. coli* expressed GST-PARG.

FIG. 11 depicts a schematic representation of the portions of the bovine PARG cDNA expressed as GST fusion constructs.

FIG. 12 depicts the cloning of the 1.8 kb PCR EcoRI fragment encoding for the 65 kDa catalytic domain of PARG.

FIG. 13 depicts an autoradiogram of an activity gel of GST-PARG fusion constructs expressed in *E. coli* and PARG expressed in baculovirus.

FIG. 14 depicts a schematic representation of the strategy used to isolate cDNA molecules encoding PARG from various organisms.

FIG. 15 depicts the domain organization of PARGs from different organisms.

FIG. 16 depicts an amino acid sequence alignment of bovine, murine, human, drosophila and *C. elegans* PARG enzymes.

FIG. 17 depicts a western blot of recombinant PARGs.

FIG. 18 depicts western blots of natural and recombinant expressed PARG.

FIG. 19 depicts the characterization of PARG by Western Blot in mouse cells of different PARP genotypes.

FIG. 20 depicts a partial restriction map of the mouse PARG locus.

FIG. 21 depicts a schematic representation of the strategy used to create PARG knockout mice.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

List of Abbreviations

ADP	adenosine diphosphate
ADPR	ADP-ribose
AMP	adenosine monophosphate
ASPCR	allele-specific PCR
bp	base pair(s)
bPARG	bovine PARG
CePARG	<i>C. elegans</i> PARG
dPARG	<i>Drosophila melanogaster</i> PARG
DTT	dithiothreitol
GSH-Sepharose	Glutathione-Sepharose 4B
GST	glutathione-S transferase
hPARG	human PARG
HPLC	high pressure liquid chromatography

-continued

ICE	interleukin-1 b converting enzyme
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase pair(s)
MDBK	Madin-Darby bovine kidney cells
mPARG	murine PARG
NAD	nicotinamide adenine dinucleotide
NLS	nuclear location signal
PADPR DHB-Sepharose	poly(ADP-ribose)-dihydroxyboronyl-Sepharose
PAGE	polyacrylamide-gel electrophoresis
PARG	poly(ADP-ribose) glycohydrolase
PARP	poly(ADP-ribose) polymerase [EC 2.4.2.30]
PCR	polymerase chain reaction
PEG-6,000	polyethylene glycol 6,000
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
PR-AMP	phosphoribosyl-adenosine monophosphate
RFLP	restriction fragment length polymorphism
SDS	sodium dodecyl sulfate
SSCP	single-strand conformation polymorphism
TPCK	Trypsin: L-1-tosylamido-2-phenylethyl chloromethyl ketone.

An “agonist” as defined herein refers to a molecule which, when bound to PARG, increases or prolongs the effect of PARG. Agonist may include proteins, nucleic acid molecules, carbohydrates, or any other molecules that bind to and modulate the effect of PARG.

An “allele” or “allelic sequence”, as defined herein refers to an alternative form of PARG. Alleles may result from at least one mutation in the nucleic acid molecule sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles, are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

An “ortholog” as defined herein refers to a nucleotide or amino acid sequence that is related to a reference nucleotide or amino acid sequence through speciation, and is therefore identical or structurally similar to the reference sequence.

A given nucleotide or amino acid sequence is said to be “substantially identical” with another sequence when the compared sequences have the same residues in the same order, excepting for any degeneracy (nucleotides) and conservative substitutions (amino acids).

A “regulatory sequence” of an expression vector is a DNA sequence necessary for inducing transcription of a gene, and includes a functional promoter and/or enhancer sequence. The term “operatively linked” as used herein means that a first nucleotide sequence, such as a regulatory element, is fused in frame with a second nucleotide sequence so as to afford a faithful transcription of the entire nucleotide sequence, which upon translation yields the desired protein.

The term “immunoreactivity” and related terms refers to the ability of antibodies and fragments thereof to bind to particular regions (antigens) presented by polypeptides and proteins, presented to the antibodies either as immunogens or targets. Typically, the binding affinity of the antibodies for their antigen is in the range 10^5 to 10^{11} , with higher affinities being preferred.

The term “specific immunoreactivity” refers to the ability of antibodies and fragments thereof to bind to particular regions (antigens) presented by polypeptides and proteins, presented to the antibodies either as immunogens or targets and not to unrelated antigens. For example, an antibody with

specific immunoreactivity to actin will bind actin but would not bind another protein, such as a polymerase, which do not share epitopes with actin.

The term “nucleic acid molecule” refers to DNA, RNA and nucleic acid molecule analogs such as PNA and the like. PNA or “Peptide Nucleic Acid” is a nucleic acid molecule analog that has a neutral “peptide-like” backbone with nucleobases that allow the molecule to hybridize to complementary RNA or DNA with higher affinity and specificity than corresponding oligonucleotides. PNA can be made to be more resistant to normal nucleases and are especially desirable, for example, in gene therapy. PNA is known to one of skill in the art and can be purchased or custom synthesized in numerous commercial laboratories including PerSeptive Biosystems, Inc. (Framingham, Ma.).

The term “modulate” means to activate or inhibit. For example, a PARG modulator may activate or inhibit PARG activity. “Modulation activity” means the amount of activation or inhibition. For example, a compound that increase PARG (or any other enzyme) activity by 10% will have a modulation activity of 10%. Conversely, a compound that decreases PARG activity by 10% will have a modulation activity of -10%.

As used herein, a given nucleotide or amino acid sequence is said to have a defined percentage of sequence similarity with another sequence when the two sequences differ by no more than the specified sequence similarity, including conservative substitutions, insertions, and deletions. Degenerate codons do not result in a change in amino acid upon translation, therefore, it is appreciated that identical amino acids can be encoded by several equivalent codons. The term “homology” and “sequence similarity” should have the same meaning for the purpose of this patent. Similarity parameters may be any generally acceptable parameter. For the purposes of this patent, percent similarity between two polymers such as nucleic acid molecules and polypeptides is preferably defined by Karlin and Altschul (11). The similarity algorithms of Karlin and Altschul are well known to those of skill in the art as exemplified by their adoption by the National Center for Biological Information. For nucleic acid molecule sequence searching, one desirable set of parameters would M (score for a pair of matching residues) at 5; N (score for mismatching residues) at -4; W (word length) at 11. For proteins, it is well known that some amino acids are similar and that substitution would be conservative. That is, for example, the replacement of an acidic amino acid with another acidic acid would be consider a conservative mismatch while the replacement of an acidic amino acid with a basic amino acid would be consider a more divergent mismatch. Preferably, the parameters for a desirable protein similarity determination are expressed in the sequence similarity matrix BLOSUM62 as described in Henikoff & Henikoff (12). Other similarity matrixes that are also preferred in the invention are PAM40, PAM120 and PAM250 as described in Altschul (13).

The rapid synthesis of ADP-ribose polymers that occurs in response to DNA strand breaks is accompanied by very rapid polymer turnover, indicating that PARP and PARG activities are closely coordinated as cells respond to DNA damage. While PARP has been widely studied, information concerning structure and function relationships of PARG is much more limited. The present invention discloses the isolation of a cDNA encoding the bovine, human, murine and drosophila PARG and their deduced amino acid sequences.

The availability of PARP cDNA has allowed a number of molecular genetic approaches to study the function(s) of

ADP-ribose polymer metabolism and the availability of PARG cDNA should allow the design of additional molecular genetic approaches for studying this metabolism. For example, disruption of the gene encoding PARG in mice containing a normal PARG gene will allow the determination of whether other cellular enzymes can replace PARG in the turnover of ADP-ribose polymers and/or whether development of animals will occur in the absence of PARG. Alternatively, disruption of the PARG gene in mice containing a disrupted PARG gene may provide insights for the coordinated function of PARP and PARG.

One embodiment of the invention is directed to a deoxyribonucleic acid (DNA) molecule that encodes a polypeptide having poly(ADP-ribose) glycohydrolase (PARG) activity. Preferably, the molecule is of mammalian origin, such as, for example, of human origin.

In a preferred embodiment, a DNA molecule of the invention comprises a nucleotide sequence with at least about 70% sequence similarity with a sequence shown in a sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9. Higher degrees of sequence similarity, such as about 80%, about 90%, and about 100% are preferred. Most preferred is a DNA molecule comprising a nucleotide sequence substantially identical with any one of sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9. It is preferred that a DNA molecule of the present invention comprises at least about 1000 nucleotides and has a nucleotide sequence with at least 80% sequence similarity with a sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9. Most preferably, the DNA molecule consist of a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9.

For a DNA molecule of the present invention based on a human PARG gene it is preferred that the molecule comprises a nucleotide sequence that shows similarity to the sequence shown in SEQ ID NO: 3 from about residue 2113 to about residue 3105. More preferably, the sequence similarity is from about residue 1240 to about residue 3105. Still more preferably, the DNA molecule comprises a nucleotide sequence similarity to the coding sequence for the full-length hPARG as shown in SEQ ID NO: 3 from about residue 175 to about residue 3105.

A DNA molecule of the present invention affords probes and primer molecules that can be used in hybridization assays and PCR amplification. An exemplary oligonucleotide is less than about 1000 residues in length and comprises a nucleotide sequence at least about 10 residues long to ensure hybridization. Preferably, the at least about 10 residue region of the oligonucleotide is complementary to a sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9. Typically, the oligonucleotide will be a DNA molecule, which can be labeled by any method as desired, for example, with a radiolabel, a fluorescence label, or chemi-luminescent label.

Another embodiment of the invention is directed to a nucleic acid molecule that hybridizes to in a nucleic acid blot (Southern blot, Northern blot) to a sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 under stringent hybridization conditions. A nucleic acid blot may be made using techniques defined in Molecular Cloning, Second Edition, Sambrook et al., Cold Spring Harbor Press, Cold Spring Harbor, N.Y. DNA to be analyzed may be separated in agarose or acrylamide gels. The DNA may be transferred to nylon or nitrocellulose membrane

using techniques known to those in the art. Stringent hybridization condition may be for example, prehybridizations 42° C. in 50% formamide, 0.25 M sodium phosphate buffer, pH 7.2, 0.25 M NaCl, 7% SDS, 1 mM EDTA for 10 hours, 100 ug denatured salmon sperm DNA, hybridization at 42° C. in 50% formamide, 0.25 M sodium phosphate buffer, 100 ug denatured salmon sperm DNA. pH 7.2, 0.25 M NaCl, 7% SDS, 1 mM EDTA, 1 ng/ml probe with a specific activity of 10⁹ cpm/ug DNA, for 16 hours. The probe may comprise any contiguous sequence from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. Preferably, said contiguous sequence is at least about 50 bases long, more preferably, the contiguous sequence is at least about 75 bases long, such as at least about 100 bases, at least about 200 bases long or at least about 300 bases long. The complete sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. Methods of labeling probes to with radioactive labels are known to those of skill in the art.

Method of washing after stringent hybridization are known. A stringent washing may comprise, for example, two washes at in 2× SSC, 0.1% SDS for 15 minutes each at room temperature; two washes in 0.2×SSC, 0.1% SDS for 15 minutes each at room temperature; and a final three washes in 0.2×SSC, 0.1% SDS for 15 minutes each at 60° C. The final wash may be increased in temperature for reduced background. For example, the final wash may be a final three washes in 0.2×SSC, 0.1% SDS for 15 minutes each at 65° C. or a final three washes in 0.2×SSC, 0.1% SDS for 15 minutes each at 68° C.

If a radioactive probe is used, hybridization may be monitored using known techniques such as autoradiogram or a two dimensional measurement of radioactivity.

An anti-sense oligonucleotide is also afforded by the present invention. The anti-sense molecule is typically less than about 1000 residues in length to ensure ease of synthesis, and hybridizes to an RNA molecule, e.g., messenger RNA, which has at least 70% sequence similarity with a sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9. Preferably, the anti-sense molecule is at least about 10 nucleotides in length to ensure hybridization with mRNA. Even more preferably, the anti-sense molecule may be at least about 15 nucleotides in length such as, for example, at least about 20 nucleotides in length; at least about 30 nucleotides in length; at least about 50 nucleotides in length; at least about 75 nucleotides in length; at least about 100 nucleotides in length; at least about 150 nucleotides in length; at least about 200 nucleotides in length; at least about 500 nucleotides in length; at least about 1000 nucleotides in length; or at least about 1500 nucleotides in length. It is also preferred that the molecule has a ribozyme activity so that it can degrade the mRNA that it binds to.

An antisense oligonucleotide may be used therapeutically to inhibit translation of mRNA encoding PARG. Synthetic antisense oligonucleotides may be produced, for example, in a commercially available oligonucleotide synthesizer. This invention provides a means to therapeutically alter levels of expression of a human or other mammalian PARG by the use of a synthetic antisense oligonucleotide drug that inhibits translation of mRNA encoding PARG. Synthetic antisense oligonucleotides, or other antisense chemical structures designed to recognize and selectively bind to mRNA, are constructed to be complementary to portions of the nucleotide sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and SEQ ID NO: 9. An antisense oligonucleotide may be designed to be stable in the

blood stream for administration to patients by injection, or in laboratory cell culture conditions, for administration to cells removed from the patient. The antisense may be designed to be capable of passing through cell membranes in order to enter the cytoplasm and nucleus of the cell by virtue of physical and chemical properties of the antisense oligonucleotide which render it capable of passing through cell membranes (e.g., by designing small, hydrophobic antisense oligonucleotide chemical structures) or by virtue of specific transport systems in the cell which recognize and transport the antisense oligonucleotide into the cell. In addition, the antisense oligonucleotide can be designed for administration only to certain selected cell populations by targeting the antisense oligonucleotide to be recognized by specific cellular uptake mechanisms which bind and take up the antisense oligonucleotide only within certain selected cell populations. For example, the antisense oligonucleotide may be designed to bind to transporter found only in a certain cell type, as discussed above. The antisense oligonucleotide may be designed to inactivate the PARG mRNA by (1) binding to the PARG mRNA and thus inducing degradation of the mRNA by intrinsic cellular mechanisms such as RNase I digestion, (2) by inhibiting translation of the mRNA target by interfering with the binding of translation-regulating factors or of ribosomes, or (3) by inclusion of other chemical structures, such as ribozyme sequences or reactive chemical groups, which either degrade or chemically modify the target mRNA. Synthetic antisense oligonucleotide drugs have been shown to be capable of the properties described above when directed against mRNA targets (14). In addition, coupling of ribozymes to antisense oligonucleotides is a promising strategy for inactivating target mRNA (15). In this manner, an antisense oligonucleotide directed to PARG may serve as a therapy to reduce PARG expression in particular target cells of a patient and in any clinical condition that may benefit from reduced expression of PARG.

It is known by those in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding PARG, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring, gene, may be produced. Thus, the invention contemplates a nucleic acid molecule that encodes a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO:6, SEQ ID NO: 8 and SEQ ID NO: 10. The invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices that would encode the oligopeptides disclosed herein. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring PARG and all such variants are to be considered as being specifically disclosed.

Although nucleic acid molecules which encode PARG and its variants preferably hybridizes under high stringency conditions to the nucleotide sequence of the naturally occurring PARG gene under appropriate conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PARG or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PARG and its derivatives and variants without altering the produced amino acid sequence include the production of RNA transcripts having more

desirable properties, such as greater half-life, than transcripts produced from the naturally occurring sequence.

In order to express a biologically active or immunologically active PARG, the nucleic acid molecule encoding PARG or functional equivalents, may be inserted into appropriate expression vector, such as, for example a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Thus, another aspect of the present invention is an expression vector comprising a regulatory sequence operatively linked to nucleic acid molecule comprising a nucleotide sequence disclosed herein. For example, an expression vector can contain a nucleotide sequence at least about 1000 base pairs in length, which has at least about 70%, about 80%, or higher, sequence similarity with a sequence shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9.

Methods that are known to those skilled in the art may be used to construct expression vectors containing sequences encoding PARG and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PARG. These include, for example, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus), insects infected with virus expression vectors (e.g., fall army worm infected with baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus; TMV) or with bacterial expression vectors (e.g., Ti or bacterial plasmids); or animal cell systems. The invention is not limited by the host cell employed.

Prokaryotic expression systems are commercially available from a number of suppliers worldwide. Prokaryotic expression vectors provide a convenient system to synthesize proteins. If it is desired to express a protein with characteristics such as immunogenic properties, 3D conformation, and other features exhibited by authentic PARG, the protein may be expressed in an eukaryotic protein expression system. The eukaryotic expression systems are numerous and include mammalian, amphibian, plant, insect, and yeast expression systems.

Yeast hosts that can be used for expression include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Hansela polymorpha*, *Kluyveromyces lactis*, and *Yarrowia lipolytica*. Yeast hosts offer the advantages of rapid growth on inexpensive minimal media and ease in large-scale production using bioreactors. Another advantage of yeast is the ability to direct expression to cytoplasmic localization or for extracellular export.

Most yeast vectors for protein expression are derivatives of the *S. cerevisiae* 2 μ (two micron) plasmid. Yeast vectors include pYES and pEST from Stratagene (La Jolla, Calif.). Constitutive gene expression by the yeast plasmid cassette can be mediated by well known promoters such as the glyceraldehyde-3-phosphate dehydrogenase promoter (TDH3); the triose phosphate isomerase promoter (TPI1); the phosphoglycerate isomerase promoter (PGK1); the alcohol dehydrogenase isozyme II (ADH2) gene promoter; GAL1 and GAL10 promoters; the metallothionein promoter from the CUP1 gene (induced by copper sulfate); and the PHO5 promoter (induced by phosphate limitation). Proper

termination of yeast transcripts is known to those in the art. Termination signals may include the MF-alpha-1, TPI1, CYC1, and PGK1 genes. These termination signals may be spliced onto the 3' end of the insert to provide proper termination.

Insect expression systems include baculovirus based vectors designed to express foreign proteins in a number of insect hosts and insect cell line hosts. Insect and insect cell lines may be of *Drosophila melanogaster*, *Aedes albopictus*, *Spodoptera frugiperda*, and *Bombyx mori* origin. Numerous expression systems comprising cells, vectors, hosts and the like can be purchased from a variety of commercial sources.

The control elements or regulatory sequences necessary for the proper expression of the insert, in this case PARG, may comprises promoters, enhancers (including both proximal and distal control elements) which interact with the host proteins to carry out transcription and translation. Such elements may vary in their strength and specificity and are known to those in the art. Depending on the vectors system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, the LacZ promoter may be used in a bacterial cell; the baculovirus polyhedrin promoter may be used in an insect cell; plant promoters such as heat shock promoters, and storage protein promoters, plant virus promoters and the like may be used in a plant cell. In a mammalian cell expression system, an SV40 promoter or EBV promoter may be used.

Methods and protocols for both prokaryotic and eukaryotic expression systems are generally known to those in the art. Further, the cells, vectors, growth medium may be purchased from commercial suppliers. The catalogs and product literature of commercial suppliers provide detailed protocols to enable the expression of proteins in prokaryotic and eukaryotic systems including bacterial, yeast, insect, insect cell, and mammalian cell systems. The product literature and catalogs of Clontech (Palo Alto, Calif.), Invitrogen (Carlsbad, Calif.), Life Technologies (Rockville, Md.), Novagen (Madison, Wis.), Plasmid (San Diego, Calif.), Quantum Biotechnologies (Montreal, Quebec, Canada), and Stratagene (La Jolla, Calif.) are incorporated herein by reference.

A further aspect of the invention is isolated proteins and protein fragments having poly(ADP-ribose) glycohydrolase (PARG) activity. Such a protein call comprise an amino acid sequence with sequence similarity of at least about 70%, about 80% or higher to a sequence shown SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10. For example, the full-length bovine PARG has a molecular weight greater than about 100 kDa, thereby distinguishing it from previously known PARGs. The protein may be purified, for example, from cell lysates using the antibodies of the invention. The purification maybe through an antibody column.

PARG polypeptides are another aspect of the invention. Polypeptides of PARG may be used, for example, to generate antibodies in an immunogenic procedure. To be effective it is preferred that the polypeptides are at least about 6 amino acid residues in length, such as for example, at least about 10 amino acids in length, at least about 20 amino acids in length, at least about 30 amino acids in length, at least about 50 amino acids in length, at least about 75 amino acids in length, at least about 100 amino acids in length, at least about 150 amino acids in length, at least about 200 amino acids in length, or at least about 400 amino acids in length. In one embodiment, the polypeptide has a molecular weight

less than about 65 kDa and with at least about 80% sequence similarity with a sequence shown in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10. The polypeptide may consist of the sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

The polypeptide of the invention may be conjugated to a larger molecule, such as, for example, keyhole limpet hemocyanin (KLH), to increase the immunogenicity of the polypeptide. The increased immunogenicity of the polypeptide will, in turn, increase the yield of antibody. Preferably, the polypeptide has a molecular weight less than about 40 kDa and with at least about 90% sequence similarity with a sequence shown in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10. The polypeptide can also be used in a wide variety of assays, e.g., as a competitor of antigen in a liquid sample in an antibody-based assay. Therefore, it is preferred that the polypeptide has poly(ADP-ribose) glycohydrolase (PARG) activity. A particularly preferred polypeptide is of human origin and comprises an amino acid sequence substantially identical with SEQ ID NO: 4 from about residue 647 to about residue 977—the C terminus catalytic region of the enzyme. Longer sequences more inclusive of the natural molecule are of course also contemplated.

The invention also encompasses PARG variants and alleles. A preferred PARG variant is one having at least 80% and more preferably at least 90% amino acid similarity to the amino acid of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10 and which retains at least one biological, immunological or other functional characteristic or activity of PARG. A most preferred PARG variant is one having at least 95%, amino sequence similarity or identity to human PARG (SEQ ID NO: 3).

Antibodies to PARG may be generated using numerous established methods that are well known in the art. One example of such a method is described in the Examples. Generated antibodies may include, for example, polyclonal, monoclonal, chimeric, single chain, Fab fragments, Fab' fragments, Fab'(2) fragments, and fragments produced by a FAB expression library. Humanized antibodies and single chain antibodies may also be produced after the amino acid sequence of effective antibodies are determined.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with PARG or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants maybe used to increase immunological response. Such adjuvants include, for example, Freund's mineral gels such as aluminum hydroxide, and surface-active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvant used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PARG have an amino acid sequence consisting of at least five amino acids and more preferably at least about 10 amino acids, such as for example about 20 amino acids or about 40 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural PARG. Short stretches of PARG amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibodies may be produced against the chimeric molecule.

Antibodies may be produced by inducing in vivo production in the lymphocyte population of a living animal or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in published procedures (16).

Antibody fragments that contain specific binding sites for PARG may be generated. For example, such fragments include the F(ab')₂ fragment, Fab fragment, Fab' fragment which can be produced by enzymatic digestion of the antibody molecule. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. (1989) *Science* 254, 1275-1281).

Therapeutic Methods

A method of preventing, treating, or ameliorating a disease condition in a patient, which disease state is affected by the level of PARG expression is also contemplated. This method entails administering a therapeutically effective amount of a poly(ADP-ribose) glycohydrolase (PARG) inhibitor or activator to the individual. Particularly, implicated disease states are neoplastic disorder, myocardial infarction, vascular stroke and neurodegenerative disorders.

In one embodiment, antisense oligonucleotides for PARG may be used alone or in combination with other chemotherapeutic agents to treat neoplastic disorder. The antisense oligo is designed to hybridize in vivo to messenger RNA expressed by the organism. The use of anti-sense molecules in a therapeutic setting is described, for example, by S. Agrawal, *Antisense Therapeutics*, Humana Press. Currently favored protocols call for the oligo to have ribozyme activity in an effort to degrade the mRNA. These methods are described, for example, in *Therapeutic Application of Ribozymes*, K. Scanlon, ed., Humana Press. Therefore, in one embodiment, an antagonist of PARG may be administered to a subject to prevent or treat neoplastic disorder.

PARG levels may be enhanced to suppress DNA repair and increase a cell's susceptibility to chemotherapy drugs. Therefore, in another embodiment, an PARG enhancer is administered to a subject along with a chemotherapeutic drug as a treatment for neoplastic disorder.

Neoplastic disorders that can be treated by PARG elevation and chemotherapy include benign and malignant neoplasm such as, for example, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy. Neoplastic disorders may include, in particular, neoplastic disorders of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. For the purposes of this invention, a neoplastic disorder is any new and abnormal growth; specifically a new growth of tissue in which the growth is uncontrolled and progressive. Malignant cancer is a subset of neoplastic disorders which show a greater degree of anaplasia and have the properties of invasion and metastasis.

The synthesis of effective anti-sense inhibitors is known. Numerous approaches have been previously described and generally involve altering the backbone of the polynucleotide to increase its stability in-vivo. Exemplary oligonucleotides and methods of synthesis are described in U.S. Pat. Nos. 5,661,134; 5,635,488; and 5,599,797 (phosphorothioate linkages), U.S. Pat. Nos. 5,587,469 and 5,459,255 (N-2 substituted purines), U.S. Pat. No. 5,539,083

(peptide nucleic acids) and U.S. Pat. Nos. 5,629,152; 5,623,070; and 5,610,289 (miscellaneous approaches). The disclosures of each of these references are incorporated herein by reference.

5 Significantly, the present invention discloses a method of identifying an agent that inhibits or activates poly(ADP-ribose) glycohydrolase (PARG) activity. Such method comprises (i) providing a liquid medium that contains a polypeptide of the present invention; (ii) contacting the polypeptide with a candidate agent, in the presence of a reference compound having affinity for the polypeptide, under predetermined assay conditions; and (iii) determining the affinity of the candidate agent for the polypeptide relative to the reference compound, thereby determining the inhibition or activation activity of the candidate agent relative to the reference compound. These determinations can be facilitated by immobilizing the polypeptide on a solid support. Alternatively, the polypeptide can be generated in vitro by culturing a cell transformed with a PARG gene under conditions effective to express the polypeptide.

Combination therapies are also afforded by the present invention in which a PARG inhibitor or activator is administered in combination with a chemotherapeutic or a "clot-busting" drug. The clot-busting drug may be, for example, tissue plasminogen activator (t-PA) or streptokinase.

In some cases it may be desired to overexpress PARG in the cells of an organism in order to achieve the correct PARG/PARG balance. In this context of gene therapy, it is desired to stably transfect target cells with a vector, such as, for example, a viral or a DNA (nucleic acid) vector, so that the desired gene is overexpressed. Gene therapy vector systems and protocols are well known and are described, for example, in the Internet Book of Gene Therapy (17) Antisense and ribozyme approaches to cancer gene therapy are described in chapters 7-9 of the *Internet Book of Gene Therapy*, and are incorporated herein by reference. Another reference is *Gene Therapy Protocols*, P. Robbins, ed., Humana Press. Furthermore, gene therapy methods have advanced greatly and are well documented in numerous issued US patents. Gene therapy may be practiced, for example, by substituting a nucleic acid molecule of the invention with the nucleic acid molecule described in the methods referred to in any issued US patents directed to Gene therapy (18).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

Diagnostic Methods

Methods of genotyping an individual for a mutant PARG allele are also afforded by the present invention. A number of protocols are available for identifying a mutant allele as described herein once the nucleotide sequence encoding PARG is known. Some exemplary methods are restriction fragment length polymorphism (RFLP), allele-specific PCR (ASPCR) and single-strand conformation polymorphism (SSCP). Armed with this information, the genetic susceptibility of an individual to an above-mentioned disease condition can be assessed.

An allele-specific method for identifying point mutations by differential PCR amplification is described by (19). A non-electrophoretic method of genotyping with allele-specific PCR employs a dye specific for double-stranded DNA (20). A method of detecting mutations referred to as single-stranded conformation polymorphism (SSCP) is pres-

ently widely employed (21). A hybrid of SSCP and Sanger dideoxy sequencing, called dideoxy fingerprinting (ddF) has recently been described (22).

Other methods of identifying allelic mutations are known to the skilled artisan. Probably the most commonly used method of genotyping is restriction fragment length polymorphism (RFLP) (23), which employs one or more restriction enzymes to identify mutant alleles occurring within a restriction site. This method has been used extensively in forensic applications and is employed commercially by such companies as Helix Biotech, Inc. Reliagene Technologies, Inc. and GenTest Laboratories, Inc. Accordingly, an instant mutant PARG allele can be detected by RFLP methods, optionally by one of these commercial entities. The above methods are most effective in the detection of homozygotes for the defective allele.

An RFLP method of identifying a mutant PARG allele in an individual entails: (i) obtaining genomic material from the individual; (ii) digesting the genomic material with a restriction enzyme having a recognition site inclusive of the mutant allele; (iii) fractionating the restriction fragments obtained from the digestion, e.g., by electrophoresis; and (iv) comparing the fractionation pattern with that obtained for a normal allele, thereby determining the presence or absence of the mutant allele.

An ASPCR method of identifying a mutant PARG allele in an individual entails: (i) hybridizing an oligonucleotide with genomic material from the individual; (ii) attempting to extend the oligonucleotide using PCR amplification; and (iii) determining the degree to which extension occurs, thereby determining the presence or absence of the mutant allele. In this method, it is preferred that the oligonucleotide hybridizes under predetermined hybridization conditions to a region immediately 5' of a predetermined mutation site in the PARG allele with the 3' terminus of the oligonucleotide complementary to an unmutated PARG allele. In these protocols, the PCR extension reaction is generally attempted at a temperature above about 50° C., more preferably above about 60° C.

A variety of protocols including ELISA, RIA and FACS for measuring PARG levels are known in the art and provide a basis for diagnosing altered or abnormal levels of PARG expression. Normal or standard values for PARG expression may be established by combining body fluids and tissue biopsies from normal mammalian subjects, rupturing the cells or permeating the cells, combining the cells with antibody under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods but preferably by photometric means. Quantities of PARG expressed in subject, control, and disease sample are compared to standard values to determine between normal, reduced or enhanced levels of PARG.

A still further aspect of the invention pertains to an antibody immunoreactive with a polypeptide of the present invention. Preferably the antibodies are specifically immunoreactive with the polypeptides of this invention such as, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10. Frequently it is desired to label the antibody, e.g., with a radiolabel, fluorescent or epitope label, to permit visualizing the antibody. Thus, antibodies immunoreactive with the PARG of this invention are afforded, which can be used to study features of PARG heterogeneity and possible modes of regulation. The high degree of sequence similarity between bovine PARC, human PARG and murine PARG permits eliciting antibodies to PARG of

one species, which are found to be cross-reactive with PARGs from other organisms. These antibodies are valuable in characterizing PARG in-vivo under defined physiological conditions in many different organisms.

Accordingly, a method of detecting a polypeptide having PARG activity, for example, a diagnostic assay, entails: (i) contacting the polypeptide with an aforementioned antibody of the invention; and (ii) determining whether the antibody immunoreacts with the polypeptide. Binding can be ascertained in a sandwich assay, as is well known, due to the ability of the antibodies to immunoreact with an epitope of PARG. Preferably, monoclonal antibodies, such as those prepared by the method of Kohler and Milstein (24) and labeled antigens effective in competing with the polypeptide, are employed. Exemplary assays are disclosed in U.S. Pat. No. 4,375,110, the disclosure of which is incorporated herein by reference.

The present invention includes immunoreactive fragments of a PARG enzyme. Immunoreactive fragments can be fragments that can elicit an immune response that recognizes a PARG enzyme. Alternatively, immunoreactive fragments can be fragments that are specifically bound by an antibody that specifically binds a PARG enzyme. Any of a variety of methods may be employed in order to identify contiguous peptide fragments of a PARG enzyme that comprise immunoreactive sequences. PARG enzymes may be fractionated by proteases, cyanogen bromide, etc. and the resultant fragments assessed for their capacity to specifically bind anti-PARG antibodies.

In an alternative embodiment, one or more synthetic peptides may be prepared in order to locate contiguous amino acid sequences that are immunoreactive. The peptides may have a sequence that includes a series of contiguous amino acids that are identical to a series of contiguous amino acids of a PARG enzyme. The peptides may be of about six amino acids to about 500 amino acids in length. The peptide may also include sequences that are not identical to sequences of a PARG so long as it includes at least about six contiguous amino acids that are identical to about six contiguous amino acids of a PARG enzyme. In a preferred embodiment, the peptide will be about 50 amino acids in length. In other preferred embodiments the length of the peptide may be from about six amino acids to about 30 amino acids.

The peptides of the present invention may comprise amino acid sequences that elicit antibodies that specifically bind to the peptide or to a PARG enzyme. Alternatively, the peptides may contain sequences that are specifically bound by anti-PARG antibodies. Peptides that are bound by anti-PARG antibodies may be identified through the use of Epitope Scanning™ strategy (Cambridge Research Biochemicals, Inc.). Thus, the linear sequence of amino acids of a particular PARG enzyme is used to construct a set of peptides of defined length which overlap other members of the set by one or more residues. The peptides may be any length; however, lengths of from about 6 to 25 amino acids are preferred. In selecting the length, a general consideration is that antibodies that recognize linear native epitopes constitute approximately 60–70% of the anti-protein antibody population (25).

The number of overlapping amino acids will generally be more than half of the length of the peptides. That is, if the peptides are about 20 amino acids long, the overlap may be 11 or more amino acids long. In preferred embodiments, each peptide will be selected such that the number of overlapping amino acid residues in adjacent peptides is from

about (n-1) to (n-3), where "n" is the number of amino acids in the peptide. An overlap of (n-1) is particularly preferred. Thus, in a particularly preferred embodiment, a first peptide may have the amino acid sequence of residues 1-10 of a PARG enzyme, a second peptide may have the amino acid sequence of residues 2-11 of the same PARG enzyme, a third peptide may have the sequence of residues 3-12 of the same PARG enzyme and so on until the entire sequence of the PARG enzyme has been synthesized in fragments.

The peptides may be synthesized using any means known to those of skill in the art. In a preferred embodiment, the peptide will be synthesized using an automated synthesizer such as a multipin peptide synthesis system. Such systems or peptides synthesis services are commercially available from suitable providers known to those skilled in the art.

To identify suitable peptides, each peptide is introduced into a well of a microtiter plate, and assayed for its ability to bind to antibodies elicited by a PARG enzyme. Such assays may be conducted in various ways known to those skilled in the art. One suitable assay is conducted by immobilizing a peptide on the surface of a well and then contacting the peptide with a solution containing an anti-PARG antibody. After washing, the well is contacted with a labeled antibody that specifically binds to the anti-PARG antibody. Thus, the presence of label in the well indicates that the anti-PARG antibody bound to the immobilized peptide. Another preferred method of determining the ability of the peptide to be specifically recognized by anti-PARG antibodies is a competitive ELISA.

Once a particular peptide has been found to bind to anti-PARG antibodies, the peptide can be used to elicit monospecific antibodies. By immunizing an experimental animal with a single peptide containing a single antigenic determinant, the antibodies elicited will all specifically bind to the same antigenic determinant even though the antibodies are not monoclonal.

Where desired, the peptides can be modified to increase their immunogenicity. Thus, they may be modified to contain an amino-terminal and/or a carboxyl-terminal cysteine or lysine residue with or without spacer arms. The peptides may be conjugated to carriers such as bovine serum albumin, ovalbumin, human serum albumin, KLH (keyhole limpet hemocyanin) or tetanus toxoid. The use of human serum albumin is preferred over ovalbumin or bovine serum.

The peptides, alone or conjugated to a carrier, may be themselves capable of eliciting an antibody response when administered to an experimental animal. Alternatively, the peptides, alone or conjugated to a carrier, may be administered in conjunction with an adjuvant. Those skilled in the art will understand that a variety of materials may function as adjuvants. Examples of possible adjuvants include, but are not limited to, Freund's complete adjuvant, Freund's incomplete adjuvant, lipopolysaccharide (LPS) and the like. Any material that increases the immune response to a fragment of a PARG enzyme may be used as an adjuvant.

The ability to produce large amounts of active PARG enzyme permits, for the first time, the large scale screening of chemical libraries for molecules capable of inhibiting or activating PARG enzymatic activity. The screening may be conducted using any assay for PARG known to those skilled in the art. In a preferred embodiment, the screen may be conducted using the TLC based assay described by Ménard, et al. (26). A known amount of PARG will be incubated under standardized conditions with [³²P]-poly(ADPR) in the presence of inhibitor or activator. After an appropriate period

of time, the reaction will be stopped and the reaction mixture separated on PEI-F cellulose TLC plates. The TLC plates may be developed in an appropriate solvent system such as methanol followed by 0.3N LiCl. The amount of ADPR released in the reaction will be quantified and the effect of the inhibitor or activator on enzymatic activity will be determined. Typical reaction conditions are 50 mM potassium phosphate (pH 7.5) at 37° C. in the presence of 25 μM [³²P]-poly(ADPR). The concentration of the inhibitor or activator can be varied as necessary to determine the K_i value of the inhibitor or activator according to standard procedures.

Another embodiment of the invention is directed to a method of altering the response of the cell to a genotoxic stress by modulating the concentration of ADPR polymers. As discussed above, the metabolism of ADPR polymers is critical in determining the fate of cells subjected to genotoxic stress. The modulation can be either an increase or a decrease in the concentration of the polymers. In one embodiment of the present invention, the concentration of ADPR polymers can be decreased by the use of a gene therapy vector expressing a high level of PARG. In another embodiment of the present invention, the concentration of polymers can be increased by inhibiting the enzymatic activity of the PARG enzyme by the addition of inhibitors or activators identified as described above. Alternatively, the concentration of ADPR polymers can be increased by interfering with the endogenous expression of PARG enzymes using antisense oligonucleotide technology.

Knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The preparation and delivery of antisense therapeutics is well known to those skilled in the art. For example, antisense therapeutics have been used to treat neoplastic disorder as exemplified by Smith, U.S. Pat. No. 5,248,671, specifically incorporated herein by reference. Additional examples of antisense therapeutics are provided by Miller, U.S. Pat. Nos. 4,511,713; and 4,757,055, specifically incorporated herein by reference.

In the present invention, an oligonucleotide having a sequence complimentary to the mRNA of the PARG gene will be prepared. Such an oligonucleotide is said to be an antisense oligonucleotide with respect to the PARG gene. The oligonucleotide may be RNA or DNA or a may contain both RNA and DNA portions. The oligonucleotide may contain modified bonds so as to enhance the stability of the oligonucleotide and render it more resistant to the action of cellular nucleases. For example, the oligonucleotide may be constructed with phosphorothioate nucleotides, phosphonate nucleotides and other types of modified nucleotides known to those skilled in the art. The structure of the oligonucleotide maybe altered so as to include other types of bonds that do not naturally occur in oligonucleotides. For example, adjacent nucleosides might be joined using linear alkyl chains, peptide bonds or other types of structures. The only limitation is that the resulting oligonucleotide remains capable of hybridizing to the target PARG mRNA.

The antisense oligonucleotides may be delivered by any means customarily used in the art. For example, the oligonucleotide may be delivered in neutral liposomes, cationic liposomes or by ballistic high speed injection. Alternatively the DNA sequence encoding, the antisense oligonucleotide may be inserted into a gene vector and the vector may be introduced into target cells. The vector may be any type of gene therapy vector known to those skilled in the art. Preferred embodiments include, plasmid vectors and viral

vectors. Viral vectors are seen to include those vectors customarily used for gene therapy applications including, but not limited to, retroviral vectors, vaccinia virus vectors, herpes virus vectors, adenovirus vectors and adeno-associated virus vectors. Upon introduction of the vector into target cells, the vector will direct expression of a nucleic acid molecule comprising the appropriate sequence to hybridize with the mRNA encoding a PARG enzyme. In a preferred embodiment, introduction of the vector into the target cell will result in the production of an RNA molecule that hybridizes with the mRNA of a PARG enzyme and also includes one or more additional RNA sequences capable of functioning as a ribozyme. The ribozyme portion of the molecule will cause the cleavage of the mRNA encoding the PARG enzyme thereby preventing the production of PARG.

Therapeutics of this type may be used to treat a wide variety of conditions. In one embodiment, an antisense therapeutic will be used to treat neoplastic disorder. In a preferred embodiment, an antisense therapeutic of the present invention will be delivered in combination with a currently known chemotherapeutic agent. In general, chemotherapeutic agents function by disrupting the integrity of DNA in target cells. Since the recovery of a cell from such DNA disruption is highly dependent upon the normal ADPR polymer metabolism, the presence of the antisense therapeutic will have the effect of chemosensitizing the neoplastic cells by disturbing the ratio PARG and PARP.

In another preferred embodiment, the antisense oligonucleotides of the present invention may be used to treat a variety of conditions caused by genotoxic oxidative stress. Examples include cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. It has been shown that inhibition of ADPR polymer synthesis provides protection against cellular damage caused by nitric oxide injury. Zhang, et al., U.S. Pat. No. 5,587,384, specifically incorporated herein by reference, teach that decreasing the amount of ADPR polymers formed can result in protection against nitric oxide induced neurotoxicity. As discussed above, decreasing the amount of ADPR polymers in the cell can be accomplished by the introduction of gene therapy vector expressing PARG, thus, the present invention can be used to treat neurodegenerative conditions resulting from oxidative stress.

CONCLUSION

The synthesis and rapid turnover of ADP-ribose polymers is an immediate cellular response to DNA damage. Reported here is the isolation and characterization of cDNAs encoding various poly(ADP-ribose) glycohydrolase (PARG) enzymes responsible for ADP-ribose polymer turnover. PARG was isolated from bovine thymus, yielding a protein of approximately 59 kDa. Based on the sequence of oligopeptides derived from the enzyme, polymerase chain reaction products and partial cDNA clones were isolated and used to construct a putative full-length cDNA. The cDNA of approximately 4.1 kb pairs predicts expression of a protein of approximately 111 kDa, nearly twice the size of the isolated protein. A single transcript of approximately 4.3 kb pairs is detected in bovine kidney poly(A)⁺ RNA, consistent with expression of a protein of 111 kDa. Expression of the cDNA in *Escherichia coli* results in an enzymatically active protein of 111 kDa and an active fragment of 59 kDa. Analysis of restriction endonuclease fragments from bovine DNA by Southern hybridization indicate that PARG is encoded by a single copy gene. Taken together, the results indicate that previous reports of multiple PARGs can be

explained by proteolysis of an 111-kDa enzyme. The deduced amino acid sequence of the bovine PARG shares little or no sequence similarity with differing types of known proteins; however, it contains a putative bipartite nuclear location signal as would be predicted for a nuclear protein. The availability of cDNA clones for PARG should facilitate structure-function studies of the enzyme and its involvement in cellular responses to genomic damage.

Other embodiments and advantages of the invention are set forth, in part, in the description that follows and, in part, will be obvious from this description and may be learned from practice of the invention.

EXAMPLES

Example 1

Purification of Bovine PARG

PARG was purified from bovine thymus tissue (Pel-Freez, Rogers, Ak.) by modifications of previously published procedures (27). The enzyme was isolated up to the polyethylene glycol (PEG)-6,000 fractionation step as described previously (28). However, DNA-agarose and heparin-Sepharose chromatographic steps used previously were omitted, and the PEG-6,000 fraction was applied directly to an affinity matrix of poly(ADP-ribose)-dihydroxyboronyl-Sepharose (PADPR DHB-Sepharose). The active fractions eluted from PADPR DHB-Sepharose (25 ml) were pooled, placed in dialysis tubing, concentrated against dry PEG-20,000 to approximately 12 ml, and dialyzed against 2 liters of 20 mM potassium phosphate buffer, pH 8.0, 0.1% Triton X-100, 5 mM β -mercaptoethanol, 0.1 mM thioglycolic acid, 0.4 M KCl (buffer A). The sample was loaded onto a 1.0x11-cm Toyopearl AF-Red (Supelco) column, and PARG was eluted with an 80-ml linear gradient of 0.4-2 M KCl in buffer A. The active fractions, eluting at approximately 1.25 M KCl, were pooled, placed in dialysis tubing, concentrated against solid sucrose to approximately 9 ml, and dialyzed against 20 mM potassium phosphate buffer, pH 7.2, 0.75 M KCl, 0.1% Triton X-100, 10% glycerol, 5 mM β -mercaptoethanol, 0.1 mM thioglycolic acid. PARG activity was determined as described by Ménard and Poirier (29), and protein content was determined by the method of Bradford (30). The final preparation was quantified by SDS-PAGE (31) and Coomassie Blue staining to compare the intensity of the protein band with a known amount of bovine serum albumin (32).

The purification procedure for the bovine thymus PARG summarized in Table 1 is typical for results obtained from six separate preparations of the enzyme. Purification from 500 g of bovine thymus achieved approximately 50,000-fold purification and yielded approximately 20 μ g of purified protein. An aliquot of the purified enzyme was precipitated with trichloroacetic acid, washed with acetone, resuspended in SDS-PAGE sample buffer, separated on a 10% SDS-PAGE gel, and stained with Coomassie Blue. Analysis of the final preparation of SDS-PAGE revealed that more than 95% of the protein migrated at an apparent molecular mass of approximately 59 kDa (FIG. 2). In FIG. 2, an aliquot of the purified enzyme was precipitated by TCA, washed with acetone, resuspended in SDS-PAGE sample buffer, separated on a 10% SDS-PAGE gel and stained with Coomassie blue. The positions of molecular weight marker proteins are shown.

TABLE 1

Purification of PARG from bovine thymus					
Step	Protein mg	Total activity units	Specific activity units/mg protein	Yield %	Purification -fold
Crude extract	27,800	57,400	2.06	100	1.0
Protamine sulfate	12,500	58,000	4.64	101	2.3
Ammonium sulfate	4,480	30,000	6.70	52	3.3
CM-Sepharose	171	19,100	112	33	55
PEG 6000	23.0	7,530	327	13	160
PADPR-DHB-Sepharose	1.30	6,730	5,180	12	2,500
Toyopearl AF-Red	0.023	2,260	98,300	4	48,000

Example 2

Peptide Sequencing

Prior to proteolytic fragmentation, the purified bPARG (40 μ g in 100 μ l of 0.4 M ammonium bicarbonate buffer, pH 8.0, 8 M urea) was incubated in a final concentration of 2.2 mM dithiothreitol at 56° C. for 15 min. Iodoacetamide was added to a final concentration of 2.0 mM, and the sample was incubated at 25° C. for 15 min. After dilution with an equal volume of water, 1.5 units of immobilized L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Pierce Chemical, Rockford, Ill.) was added, and the sample was incubated at 37° C. for 18 h with gentle rotary shaking. Finally, the mixture was subjected to centrifugation at 16,000 \times g for 5 min to separate the tryptic fragments from the immobilized trypsin. The tryptic fragments were adjusted to 0.05% in trifluoroacetic acid and separated on a 4.6 mm \times 25 cm, Microsorb MV, C₄ reversed-phase HPLC column (Rainin) eluted with an 80-min linear gradient from 4 to 44% acetonitrile in 0.05% trifluoroacetic acid. Four oligopeptide fractions, with approximate elution times of 61, 63, 68, and 75 min, were selected for peptide sequence analysis the Edman degradation method. Amino acid sequence data of four oligopeptides, designated by their approximate HPLC elution times from the reversed-phase column, are shown in Table II.

TABLE II

Amino acid sequence of oligopeptides derived from bPARG				
Oligopeptide	Amino Acid Sequence			SEQ ID NO:
10	20	30		
68 LFTEVLHDNE	CLIIITGTEQY	SEYTYGAETY	R	SEQ ID NO: 11
63 AYCGFLRPGV	SSENL SAVAT	GNXGCGAFG		SEQ ID NO: 12
61 FLINPELIVS	R			SEQ ID NO: 13
75 IALXLPNIXT	QPIPLL			SEQ ID NO: 14

Example 3

cDNA Cloning

To obtain cDNA clones encoding bovine PARG, PCR amplification experiments were followed by the screening of two different bovine cDNA libraries. FIG. 3 depicts the alignment of the DNA sequences of two PCR products and eight λ gt11 cDNA clones used to identify the cDNA coding

for bovine PARG. The two PCR products and clones 1 and 2 were obtained from the bovine thymus cDNA library. Clones 3–8 were obtained from the bovine kidney cDNA library. The positions of restriction sites used in this study are shown, and the top diagram shows the consensus clone, denoting the relative location of the coding regions for oligopeptides, 75, 61, 68, and 63 as well as the open reading frame and non-coding regions. For each of the cDNA inserts characterized, the sequence of both strands was determined by the dideoxynucleotide chain termination method using Sequenase™ (U.S. Biochemical Corp., Cleveland, Ohio).

The first step leading to the isolation of cDNA clones was to synthesize two multi-degenerate 17-mer primers, GAY-CAYAAAYGARTGYT (SEQ ID NO: 15) and CKRTANG-TYTCNGCRTA (SEQ ID NO: 16) (where Y represents T/C, R is A/G, K is T/G, and N is A/T/C/G), based on two regions of the SEQ ID NO: 11; "DHNECL" (amino acids 7 to 12 of SEQ ID NO: 11) and "YAETYR" (amino acid 26 to amino acid 31 of SEQ ID NO: 11) (Table II). Using the multidegenerate primers and an oligo(dT)-primed bovine thymus cDNA λ gt11 library BL1019b from Clontech (Palo Alto, Calif.), PCR amplification generated a 74-bp DNA fragment with a deduced amino acid sequence identical to the corresponding region of oligopeptide 68. Next, two specific 24-mer oligonucleotide primers, ATCATCACAGGTACT-GAGCAGTAC (SEQ ID NO: 17) and GCCTGTGTAT-TCACTGTACTGCTC (SEQ ID NO: 18), based on the sequence of this 74-bp DNA were used in combination with λ gt11 forward and reverse primers to amplify PCR products 1 and 2 from the bovine thymus library. PCR product 12 contained 231 bp of sequence including the region encoding the N-terminal region of oligopeptide 68 (SEQ ID NO: 11) and the entire sequence of oligopeptide 61 (SEQ ID NO: 13). PCR product 2 contained 757 bp, which included a sequence encoding the C-terminal region of oligopeptide 68 (SEQ ID NO: 11) and the entire sequence of oligopeptide 63 (SEQ ID NO: 12).

The sequence information obtained from PCR products 1 and 2 was used to isolate cDNA clones obtained by the screening of bovine thymus and bovine kidney cDNA libraries. A 518-bp EcoRI-HindIII fragment from PCR product 2 was used as a probe to screen approximately 1 \times 10⁶ independent clones from the bovine thymus library. Two positive cDNA clones (clones 1 and 2) were isolated, which overlapped PCR products 1 and 2. However, attempts to obtain clones from the bovine thymus library that contained sequence 5' to clone 2 were unsuccessful. Thus, a 231-bp EcoRI-KpnI fragment from clone 2 was used as a probe to screen approximately 5 \times 10⁵ independent clones of the bovine kidney 5' stretch plus cDNA λ gt11 library BL3001b (Clontech, Palo Alto, Calif.). Three positive cDNA clones (clones 3–5) were obtained, all of which contained sequence 5' to clone 2. Each of these clones also contained a sequence encoding oligopeptide 75. Clones 1–5 provided multiple overlapping sequences in the 3'-terminal portion of a consensus cDNA, but additional clones were sought to obtain overlapping sequences for the 5'-terminal region. Thus, a 436-bp EcoRI-KpnI fragment located at the 5' end of clone 3 was used as a probe to screen approximately 6 \times 10⁵ independent clones of the bovine kidney library. Clones 6–8 provided overlapping sequences for the 5'-terminal region. The full-length cDNA was constructed by ligating a 3.9-kb XbaI-NsiI fragment from pWL11 (clone 1 cDNA insert in pTZ18R (33)) and a 3.0-kb NsiI-XbaI fragment from pWL13 (clone 4 cDNA insert in pTZ18R). The resulting plasmid, termed pWL30, contained the 4,070-bp full-length cDNA.

FIG. 3 shows an alignment of the DNA sequences of two PCR products and eight λ gt11 cDNA clones used to identify the cDNA coding for bovine PARG. The two PCR products and Clones 1 and 2 were obtained from the bovine thymus cDNA library. Clones 3 through 8 were obtained from the bovine kidney cDNA library. The position of restriction sites used in this study is shown and the top diagram shows the consensus clone, denoting the relative location of the coding regions for oligopeptides 75, 61, 68, and 63 as well as the open reading frame and non coding regions.

The nucleotide sequence of cDNA coding for bovine PARG is shown in the sequence listing as SEQ ID NO: 1. The deduced amino acid sequence of the enzyme is shown in the sequence listing as SEQ ID NO: 2. The four oligopeptides sequenced from purified enzyme is within SEQ ID NO: 2. They are IALCLPNICTQPIPLK (amino acid 601 to 617, SEQ ID NO: 2); LINPELIVSR (amino acid 761 to 770, SEQ ID NO: 2); LFTEVLVDHNECLITGTEQYSEYTGAYETYSR (amino acid 771 to 801, SEQ ID NO: 2) and AYCGFLRPGV PSENLSAVAT GNWGCFAFGGDAR (amino acid 849 to 880, SEQ ID NO: 2). The combined nucleotide sequence of Clones 1 through 8 predicted a full-length cDNA clone of 4,070 bp containing 257 bp of 5'-non-coding sequence, a single open reading frame of 2,931 bp (beginning at the ATG at position 258 of SEQ ID NO: 1) and a 3'-non-coding region of 882 bp. and the deduced amino acid sequence which predicts a protein of 977 amino acids and a molecular weight of 110.8 kDa.

Example 4

Analysis of the Sequence of Bovine PARG

The cDNA clone (SEQ ID NO: 1) has features typical of cDNAs that code for mammalian proteins. These include (i) an oligo A (putative poly(A)+) sequence at the 3'-end, (ii) a polyadenylation signal (AAATAAA) 12 bp upstream from the oligo A sequence, (iii) a sequence of ATTTA in the 3'-untranslated region thought to play a role in selective mRNA degradation in mammalian cells (34), (iv) a single open reading frame, and (v) a nucleotide sequence around the first start codon commonly found at known sites of initiation of translation (35). The evidence that the cDNA clone constructed represents a full-length or nearly full-length clone for PARG is shown by the observation that hybridization of poly(A)+ RNA from bovine kidney cells with the cDNA showed a single band of hybridization of approximately the same size as the cDNA under stringent hybridization conditions (set forth above) (FIG. 4).

The nucleotide sequence encoding bovine PARG indicates that PARG shares little or no sequence similarity with other known sequences. A search of sequence data banks has failed to reveal significant sequence similarity with any sequences coding for known proteins. A strong sequence similarity has been observed with human and rat cDNA clones that likely represent partial clones for PARG from these species. Examination of protein sequence databases such as Genbank and SwissPro also has shown that the deduced amino acid sequence of PARG lacks any sequence similarity with known proteins. However, the amino acid sequence shares a significant similarity with a protein sequence from *Caenorhabditis elegans* that may represent the PARG protein from this organism (36).

The deduced amino acid sequence of PARG has been examined for a number of structural motifs that can be predicted from the primary amino acid sequence. The expressed PARG protein was observed to be able to form dimers stable to SDS-PAGE conditions. In that regard, residues 871-907 show significant homologies to known leucine zipper dimerization sequences (37).

Another motif identified is a putative bipartite nuclear location signal (NLS) (38). It is interesting that PARG also contains a bipartite NLS (39). FIG. 5 compares deduced amino acid sequences in the NLS region of the bovine PARG, and regions of putative PARG sequences from human, mouse and *C. elegans*, with the NLS region of PARG from seven different organisms. Conserved residues are noted in bold and the amino acid distances are from the amino terminal methionine residue. Abbreviations and references for the sequences shown are as follows: bPARG, bovine PARG (SEQ ID NO: 19); hPARG, human PARG (SEQ ID NO: 20); mPARG, murine PARG (SEQ ID NO: 21); CePARG, *Caenorhabditis elegans* PARG (SEQ ID NO: 22); hPARG, human PARG (SEQ ID NO: 23; 40); mPARG, murine PARG (SEQ ID NO: 24; 41); bPARG, bovine PARG (SEQ ID NO: 25, 42); aPARG, chicken PARG (SEQ ID NO: 26; 43); X1PARG, *Xenopus laevis* PARG (SEQ ID NO: 27; 44); DmPARG, *Drosophila melanogaster* PARG (SEQ ID NO: 28; 45); SpPARG, *Sarcophaga peregrina* PARG (SEQ ID NO: 29; 46). In FIG. 5, conserved residues are noted in boldface type, and the amino acid distances are from the amino-terminal methionine residue. Sequence alignment of putative bipartite nuclear localization signal of bovine, human and murine PARG compared to the nuclear localization signal of PARG from different organisms. The putative NLS of PARG fulfills the criteria for bipartite NLS in that it contains conserved acidic and basic amino acid residues at two different locations each within the region of sequence similarity to the NLS of PARG (47).

A surprising finding was that the bovine PARG cDNA clone codes for a protein of approximately 111 kDa, which is nearly twice the size of the PARG protein isolated from bovine thymus (FIG. 2). It indicates that PARG contains a protease sensitive site that, following proteolysis, yields a protein fragment of approximately 59 kDa that still retains enzymatic activity. Several pieces of evidence favor this possibility. (i) Expression of the carboxyl terminal portion of the cDNA resulted in enzymatic activity (FIG. 6, bar 5). (ii) All of the oligopeptides sequenced were located in the carboxyl terminal half of the protein (FIG. 3, FIG. 6 and Table 2). (iii) The only protein, other than 59 kDa protein detected in the thymus preparation was approximately 111 kD (FIG. 2). (iv) The PARG activity expressed in bacteria was sensitive to proteolysis, yielding a protein of approximately 56 kD (FIG. 6). (v) The cleavage site in PARG is in the region of the putative NLS and the PARG NLS is located in a protease sensitive site (48). Taken together with the data suggesting that bovine PARG appears to be coded for by a single copy gene (FIG. 7), proteolysis seems likely to explain the presence of PARG activity of molecular weight of approximately 74 kDa and 59 kDa in bovine thymus preparations (49). Likewise, a similar mechanism could explain previous reports of a PARG of 74 kDa isolated from nuclear fractions of guinea pig liver and human placenta (50) and a PARG of 59 kDa isolated from postnuclear fractions of guinea pig liver (51).

While proteolysis of a larger protein to yield smaller proteins retaining PARG activity seems likely to explain the size heterogeneity of PARG previously reported, it remains to be determined if proteolysis normally occurs in vivo or whether it occurs during purification of the enzyme. While the results presented here show that a full-length protein can be expressed containing PARG activity (FIG. 8), the molecular size of PARG in vivo also remains to be determined. If PARG occurs as a larger protein, an interesting possibility is that the amino terminal region may be involved in the regulation of enzymatic activity.

Example 5

Expression of bPARG in *Escherichia coli*

To determine whether the isolated cDNA encoded PARG, bPARG was expressed using two different bacterial expression systems, the pTrcHis Xpress System™ (Invitrogen, Carlsbad, Calif.), in which the expressed protein contains a leader polyhistidine sequence, and the glutathione S-transferase (GST) gene fusion system (Pharmacia Biotech Inc., Piscataway, N.J.). For expression in the pTrcHis Xpress system, three different DNA fragments were amplified and inserted into the pTrcHis expression plasmid. Constructs A and B contained the entire opening reading frame of 110.8 kDa, which together with the fusion partner predicted a protein of about 115 kDa. Construct B also contained the 3'-untranslated region of the clone. Construct A, containing the cDNA sequence-3 to 2,946, was prepared by subcloning a 2.9kb XhoI-EcoRI DNA fragment amplified from pWL30 with primers WIN34 (GCTGCGGGTCTCGACGATGAGTGC GGGC) (SEQ ID NO: 30) and WIN15 (GCGTCTAGAAITCACTTGGCTCCTCAGGC) (SEQ ID NO: 31). Construct B, containing the cDNA sequence-3 to 3,813, was prepared by subcloning a 2.8-kb XhoI-EcoRI DNA fragment amplified from pWL30 with primers WIN34 (SEQ ID NO: 30) and WIN33 (CCGGAATTCGGGTTTTTTGTTAATGAAAATTTATTAAC) (SEQ ID NO: 32). Construct C, containing cDNA sequence 964–2,946, was prepared by subcloning a 2.0-kb DNA fragment amplified from pWL13 with primers WIN14 (TCAGAGCAGATGAACTCGAGCAGTCCAGG) (SEQ ID NO: 33) and WIN15 (SEQ ID NO: 31). Since the isolated PARG of approximately 59 kDa contained enzymatic activity, construct C contained only the 75-kDa carboxyl-terminal region of the PARG, which predicted a fusion protein of approximately 79 kDa.

For expression experiments of bPARG as a GST fusion protein, an insert containing the cDNA sequence from position 1138 to 2946 was prepared by subcloning a 1.8-kb EcoRI-EcoRI fragment amplified from pWL30 with the oligonucleotide CCAATTTGAAGGAGGAATCCCGC-CGCCACCATGAATGATGTGAATGC-CAAACGACCTGGA (SEQ ID NO: 34) and WIN15 (SEQ ID NO: 31) as primers. The resulting DNA fragment was inserted into the EcoRI site of the pGEX-2T expression vector, and the plasmid was used to transform *E. coli* NM522 cells.

For expression experiments, bacterial cultures were grown at 37° C. in 1% Bacto-tryptone, 0.5% yeast extract, and 0.5% NaCl to a density of approximately 0.6 A₆₀₀/ml and were induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG). Cells were collected by centrifugation, and crude extracts were prepared by sonication (10 A₆₀₀/ml) in 10 mM sodium phosphate buffer, pH 7.2, 150 mM NaCl, 0.5 mg/ml lysozyme, 0.1 mg/ml phenylmethylsulfonyl fluoride, 1 mM EDTA, 0.7 μg/ml pepstatin A, 0.5 μg/ml leupeptin, and 1 μg/ml aprotinin. Cell extracts were subjected to centrifugation, and the supernatant fraction was used for assay. PARG assay conditions were as described previously (52). Following incubations, portions of reaction mixture were analyzed by thin layer chromatography or subjected to anion exchange HPLC.

Using a thin layer chromatography assay that measures release of [³²P]ADP-ribose from [³²P]ADP-ribose polymers (53), PARG activity was detected in extracts from cells transformed by each of the constructs. FIG. 6 shows results obtained with constructs B and C. Reaction mixtures contained approximately 15,000 cpm of [³²P]ADP-ribose

polymers, and the cpm shown represent ADP-ribose released from the ADP-ribose polymers. Bar 1, a strain transformed by pTrcHis without an insert but induced with 1 mM IPTG for 5 h at 37° C. A strain containing construct B is shown without the addition of IPTG (bar 2) or after the addition of 1 mM IPTG for 1.5 h (bar 3) or 5 h (bar 4). A strain containing construct C 5 h after induction by IPTG is shown in the absence (bar 5) and presence (bar 6) of 167 μM ADP-hydroxymethylpyrrolidine diol (54). No activity was detected in cells transformed with the empty vector, but activity was detectable without induction by IPTG, indicating a leaky lac promoter. The addition of IPTG resulted in a time-dependent increase of up to approximately 4.5-fold in enzymatic activity. FIG. 6 also shows that the enzymatic activity was strongly inhibited by the presence of ADP-hydroxymethylpyrrolidine diol, a specific inhibitor of PARG (55).

In FIG. 9, material released from ADP-ribose polymers by anion exchange HPLC was analyzed. Extracts from a strain containing construct B were incubated with [³²P]ADP-ribose polymers (56), and a portion was analyzed by anion exchange HPLC as described. The elution times for AMP, ADPR, and PR-AMP are indicated by arrows. The material analyzed was PARG expressed in *E. coli*. The results indicated that the material released from ADP-ribose polymers is exclusively ADP-ribose by strong anion exchange HPLC (FIG. 9), demonstrating that the cell extracts did not contain any other ADP-ribose polymer-degrading enzymes such as phosphodiesterase, which catalyzes the formation of AMP and phosphoribosyl-AMP (57).

Anion exchange HPLC utilized a Whatman Partisil SAX column equilibrated with 7 mM potassium phosphate buffer pH 4.0, at a flow rate of 1 ml/min. The sample was diluted in the same buffer, applied to the column, and eluted with a 30-min linear gradient from 7 mM potassium phosphate buffer, pH 4.0 to 250 mM potassium phosphate buffer, 0.5 M KCl, pH 4.0.

To determine the size of the expressed enzymatic activity, an activity gel assay (58) was used. Activity gel assays for bPARG were done by casting polyacrylamide gels with automodified PARG containing [³²P]ADP-ribose polymers as described previously (59). Following electrophoresis, PARG was renatured by incubating the gels at 25° C. in 5 volumes of 50 mM sodium phosphate buffer, pH 7.5, 50 mM NaCl, 10% glycerol, 1% Triton X-100, 10 mM β-mercaptoethanol, changing the buffer every 3 h for a total of five changes. After an additional incubation at 37° C. for 3 h, gels were dried, and PARG activity was detected following autoradiography as a clear band on a black background. Cell extracts containing PARG fused to GST were examined for binding to glutathione-Sepharose 4B (GSH-Sepharose) (Pharmacia Biotech Inc.) according to the specifications of the manufacturer. No bands were produced from extracts from the IPTG-induced pTrcHisB vector that did not contain an insert. Extracts from cells transformed with a construct containing a PARG insert showed bands at approximately 115 and 59 kDa (FIG. 8). During storage at 4° C., cell extracts lost activity migrating at the higher molecular weight, while the activity at approximately 59 kDa increased.

Expression of bPARG in the pTrcHisB expression vector did not result in detectable amounts of protein by staining the Coomassie Blue. Thus, another construction was designed to overexpress a 69-kDa carboxyl-terminal region of the PARG as a fusion with GST, which allows convenient protein purification by affinity chromatography on a GSH-Sepharose column. Two hours after induction with IPTG,

strong expression of a protein migrating at approximately 90 kDa was observed. This protein bound to GSH-Sepharose and was eluted by GSH. The construct contained a thrombin cleavage site between the GST and the 69-kDa region of PARG, and the treatment of the material bound to GSH-Sepharose with thrombin resulted in the release of a protein that migrated at approximately 59 kDa. This result suggests that the protein purified from the bovine thymus may be larger than suggested by its migration on SDS-PAGE. The result of this experiment is presented in FIG. 10. Lane 1 shows extract from uninduced cells; lane 2 shows extract from cells induced with 1 mM IPTG for 2 hours; lane 3 shows proteins in extracts from cells shown in lane 2 that bound to GSH-Sepharose; lane 4 shows material released from GSH-Sepharose by treatment with thrombin.

In addition to the GST fusion construct described above, several other GST fusion proteins have been made. FIG. 11 shows the portions of the bovine PARG gene that have been expressed. The top line represents the structure of bovine PARG mRNA containing the open reading frame encoding the 111 kDa PARG protein. The different parts of PARG that have been cloned in expression vectors are represented with the size of the resulting expressed recombinant proteins. The expression of the 65 kDa catalytic domain of PARG (starting at the amino acid MNDV) in pGEX-2T as a fusion protein with glutathione-S-transferase (29 kDa) is detailed. Among the constructs, only the clone designed to express a protein of 69 kDa starting at amino acid +380 from the sequence of bovine PARG (bPARG_{MNDV}) allowed high level expression as a fusion protein with glutathione-S transferase (GST). A 1.8 kb PCR EcoRI fragment encoding for the 65 kDa catalytic domain of PARG was cloned into the EcoRI site of pGEX-2T giving pGEX-2T-bPARG_{MNDV}. This construction results in the expression of a fused polypeptide consisting of the sequence of GST. Amino acids derived from the polylinker and thrombin site and the 65 kDa domain (FIG. 12).

In addition to various constructs designed to express PARG in *E. coli*, a recombinant baculovirus expressing a functional PARG has been constructed using the methodology of Summers and Smith as set out in U.S. Pat. No. 4,879,936 which is specifically incorporated herein by reference.

bPARG_{MNDV} was cloned in baculovirus transfer vector pVL1393 using the EcoRI site. The recombinant vector was constructed as follows. An insert containing the cDNA sequence from position 1138–2946 of bovine PARG was prepared by subcloning a 1.8 kb EcoRI fragment amplified from pWL30 using oligonucleotides CCAATTTGAAG-GAGGAATTCCTCCGCGCCACCATGAAT-GATGTGAATGCCAAACGACCTGGA (SEQ ID NO: 34) and GCGTCTAGAATTCACCTGGCTCCTCAGGC (SEQ ID NO: 31, WIN15). The resulting fragment was inserted into the EcoRI site of the pVL1393 baculovirus transfer vector. The amplification introduced a Kozak consensus sequence (gaattccgcgcaccATGAA SEQ ID NO: 35) at the start site of translation to enhance expression of the recombinant protein. The resulting recombinant plasmid was cotransfected with linearize Baculogold™ baculovirus DNA (Pharming, San Diego, Calif.) into SF9 cells according to the manufacturers instructions. Recombinant viruses isolated using standard techniques. Overexpression of the recombinant protein was confirmed by Western blot and the results displayed in FIG. 13 demonstrate that the 65 kDa domain expressed in *E. coli* contained enzymatic activity (lane 2) migrating with the same apparent molecular weight as the enzyme purified from bovine thymus (lane 1).

Likewise, a construct expressing bPARG_{MNDV} domain in SF9 insect cells infected with recombinant baculovirus showed activity (lane 4) migrating with the same apparent molecular weight.

Example 6

Northern Blot Analysis

An surprising feature of the consensus full-length cDNA clone was that it predicted expression of a protein of approximately 111 kDa (FIG. 3, SEQ ID NO: 1, and SEQ ID NO: 2), while the enzymatically active PARG from thymus had a molecular weight of approximately 59 kDa (FIG. 2). To determine the size of the RNA transcript for PARG, total RNA and poly(A)+RNA were isolated from bovine kidney (MDBK) cells and annealed using Clone 4 as the hybridization probe.

Total cytoplasmic RNA and poly(A)+RNA were isolated from bovine kidney MDBK cells (ATCC #CCL22) using TRIzol reagent (Gibco/BRL) following the manufacturer's recommendations. After the RNA was fractionated, it was then transferred to nylon membranes and hybridized with Clone 4 (FIG. 3) radiolabeled by a random hexamer priming method (21). The results are presented in FIG. 4. Total RNA (5 µg, lanes 1A and 1B) and poly(A)+RNA (4 µg, lanes 2A and 2B) were separated on a denaturing agarose gel (60). Panel A shows the ethidium bromide stained gel and panel B shows the autoradiogram of a Northern blot analysis using a random primed, ³²P-labeled DNA probe constructed from Clone 4 (FIG. 3). A single transcript of approximately 4.3 kb was detected in the poly(A)+RNA (FIG. 4, lane 2). Thus, the transcript size was consistent with the expression of a 111 kDa PARG protein.

Example 7

Southern Blot Analysis of PARG Genomic Complexity

Previous studies have reported that PARG isolated from nuclear fractions had a molecular weight of approximately 75 kDa (61), while PARG isolated from whole cell homogenates or postnuclear supernatant fractions had a molecular weight of approximately 59 kDa (62). These results suggest that either two or more genes may code for PARG or that proteolysis generates lower molecular weight forms from higher molecular weight forms. The cDNA isolated encoded a protein considerably larger than any PARG proteins previously described, consistent with the possibility that the different forms of PARG are derived from a single form by proteolytic cleavage. To test the hypothesis that PARG is encoded by a single copy gene, the genomic complexity of the PARG gene was analyzed by a Southern hybridization experiment.

Total genomic DNA was prepared from bovine thymus tissue as described previously (63) and DNA (10 µg) was digested with EcoRI, BglIII, XbaI or PstI, fractionated on a 1% agarose gel, transferred to a nylon membrane (Hybond N+, Amersham), and hybridized using an 828 bp HindIII fragment of Clone 1 radiolabeled as described for clone 4 above (64). Pre-hybridizations and hybridizations were carried out at 42° C. in 50% formamide, 0.25 M sodium phosphate buffer, pH 7.2, 0.25 M NaCl, 7% SDS, 1 mM EDTA. The blot was annealed with a ³²P-labeled DNA probe corresponding to the carboxyl terminal region of the PARG protein.

The results of the Southern blot analysis are presented in FIG. 7. Genomic DNA was digested with four different restriction enzymes, EcoRI (lane 1), BglIII (lane 2), XbaI (lane 3) and PstI (lane 4), none of which cleave within the carboxyl terminal region of the PARG cDNA. Following electrophoresis, the restriction digests were subjected to

hybridization with a probe that corresponded to the carboxyl terminal region of the PARG cDNA. The analysis displayed in FIG. 7 shows that, in each restriction digest, the probe hybridized primarily with a single restriction fragment. The fainter signals likely reflect the presence of introns in the PARC gene. This result indicates that PARG is encoded by a single copy gene in the bovine genome.

Example 8

Isolation and Characterization of PARGs From Other Species

The isolation and characterization of bovine cDNA encoding poly(ADP-ribose) glycohydrolase (PARG) has been described above. Using the information provided by the sequencing of bovine PARG, various tools were used, including public sequence databases searches and screening of cDNA libraries using PARG specific probes, to clone and sequence the cDNA and determine the primary structure of PARG from human, mouse, *Drosophila* and *Caenorhabditis elegans*. Mammalian sequences newly obtained using this combined strategy show high sequence similarity to bovine PARG (bPARG), whereas the sequences of *Drosophila* and *C. elegans* only display significant homologies in the region responsible of the catalytic activity of the protein.

The strategy followed to obtain cDNAs coding for proteins with sequence similarity to bovine PARG is summarized in FIG. 14. dBEST, GenBank, SwissProt and PIR databases were searched for PARG like sequences at the nucleotide or amino acid level using the programs BLASTn, TBLASTn (Altschul et al., 1990) respectively, available at the NIH site on the Worldwide Web, and also included in the sequence analysis package from the Genetic Computer Group, Inc. (GCG) (Madison, Wis.), version 9.1. Both programs perform pair-wise sequence comparisons on multiple nucleotide or amino acid sequences. PARG multiple sequence comparisons obtained with these programs are very similar. Box-shading of the amino acids in the multi-sequence alignment was obtained using the program BOX-SHADE (K. Hofmann and M. D. Baron). The first step involved extensive searching for sequences with bPARG similarity in various databases. As a result of this search several partial nucleotide sequences sharing extensive homologies with bPARG cDNA were obtained from the dBEST database (65). These sequences were the result of random cloning and sequencing of partial cDNAs clones obtained from mRNAs expressed in various tissues and organisms. Among them, partial cDNAs coding for PARG from human and mouse were available. One of these human clones was particularly interesting as its sequence (2500 bp long) overlapped the coding sequence of bovine PARG from aa470 to aa977 (Carboxy terminus end) and contained all the 3' untranslated region of the human PARG cDNA. This clone (No. 50859; GenBank accession number: H 17209) was requested and freely obtained from the IMAGE Consortium (in collaboration with Washington University School of Medicine in St. Louis, Mo. and Merck & Co., info@image.llnl.gov). The sequence of the clone was then completed. This partial cDNA permitted design of a radio-labeled probe (fragment HindIII-KpnI of 677 bp) specific to human PARG (SEQ ID NO: 36).

Example 9

Cloning and Sequencing

The cloning procedures used in this work generally known and are also described in details in the book, *Molecular Cloning: A Laboratory Manual* (Maniatis et al., 1982). DNA sequencing was performed using the dideoxynucleotide method of Sanger (Sanger et al., 1977). Chemical

reagents were purchased from Sigma (St. Louis, Mo.). Restriction enzymes, T4 DNA ligase were from New England Biolabs, Inc. (Beverly, Mass.), T7 DNA polymerase Sequenase from US Biochemical (Cleveland, Ohio), Calf Intestine Phosphatase from Boehringer, Mannheim (Indianapolis, Ind.). The phagemid pTZ18/19R is from Pharmacia (Piscataway, N.J.). The labeled nucleotides α -[35 S]-dATP and α -[32 P]-dCTP were purchased from ICN (Costa Mesa, Calif.). Human thymus and murine liver 5'-stretch cDNA libraries cloned in the vector λ gt 10 were from Clontech (Palo Alto, Calif.).

A single, isolated colony of C600Hfl *E. coli* strain was picked and grown in 5 ml of Luria-Bertani medium (LB)+10 mM MgSO₄+0.2% maltose overnight at 37° C. in a shaker. The bovine library lysate was diluted 1:250,000 and incubated with the C600Hfl bacterial overnight culture and 1xlambda dilution buffer. Next, LB soft top agar+10 mM MgSO₄ was added, and the entire mixture was quickly poured onto 90 mm LB agar+10 mM MgSO₄ plates. The plates were cooled briefly at room temperature to allow the inoculum to soak into the agar before they were incubated at 37° C. for 6-7 hr. The number of clear plaques was counted to determine the titer.

Plaques containing the entire library that had been plated were transferred to nitrocellulose or nylon membranes. The filters were then washed in a 1.5 M NaCl/0.5 M NaOH solution to lyse the cells. This was followed by a 5 min wash in neutralizing solution (1.5 M NaCl/1 M Tris buffer pH 8). Finally, the filters were rinsed in 0.2xSSPE (30 mM NaCl/2 mM sodium phosphate buffer pH 7.2/0.2 mM EDTA) (Sambrook et al., 1992). The filters were then dried and baked in a 80° C. oven for 2 hr to fix the lysed plaques onto the filters.

Radioactive probes were prepared using a random hexamer priming method. Pre-hybridizations and hybridizations were carried out at 42° C. in 50% formamide, 0.25 M sodium phosphate buffer, pH 7.2, 0.25 M NaCl, 7% SDS, 1 mM EDTA.

Example 10

Specific Methods Used For Library Screening

All the cloning procedures used in obtaining the additional PARG cDNAs and determining their sequences were performed essentially as described for the bovine PARG cDNA and sequence. Human thymus and murine liver 5'-stretch cDNA libraries cloned in the vector λ gt 10 were from Clontech (Palo Alto, Calif.).

Library plating and titering: A single, isolated colony of C600Hfl *E. coli* strain was picked and grown in 5 ml of Luria-Bertani medium (LB)+10 mM MgSO₄+0.2% maltose overnight at 37° C. in a shaker. The library lysate was diluted 1:250000 and incubated with the C600Hfl bacterial overnight culture and 1xlambda dilution buffer. Next, LB soft top agar+10 mM MgSO₄ was added, and the entire mixture was quickly poured onto 90 mm LB agar+10 mM MgSO₄ plates. The plates were cooled briefly at room temperature to allow the inoculum to soak into the agar before they were incubated at 37° C. for 6-7 hr. The number of clear plaques was counted to determine the titer.

Plaque lifts: Plaques containing the entire library that have been plated are transferred to nitrocellulose or nylon membranes. The filters are then washed in a 1.5 M NaCl/0.5 M NaOH solution to lyse the cells. This is followed by a 5 min wash in neutralizing solution (1.5 M NaCl/1 M Tris buffer pH 8). Finally, the filters are rinsed in 0.2xSSPE (30 mM NaCl/2 mM sodium phosphate buffer pH 7.2/0.2 mM EDTA) (66). The filters are then dried and baked in a 80° C. oven for 2 hr to fix the lysed plaques onto the filters.

Making a radioactive probe and Hybridizations: Radioactive probes were prepared using a random hexamer priming method. Pre-hybridizations and hybridizations were carried out at 42° C. in 50% formamide, 0.25 M sodium phosphate buffer, pH 7.2., 0.25 M NaCl, 7% SDS, 1 mM EDTA. This partial cDNA allowed to design a radiolabeled probe (fragment HindIII-KpnI of 750 bp long) specific to human PARG.

Example 11

Screening of a Human Thymus 5'-stretch cDNA Libraries

Multiple screenings of a human thymus 5'-stretch cDNA library were performed to complete the cloning of human PARG cDNA. For each screening a new probe was designed and used to screen approximately one million recombinants of the library. During each round of screening, overlapping clones were isolated at high stringency conditions and subcloned into the EcoRI site of pTZ18/19R phagemid using standard techniques. The different positive clones (J5, C, E1, E2, M, M', M", P, P", Of, O2) were characterized by restriction analysis, subcloned into the appropriate restriction sites of pTZ18/19R as necessary and sequenced in both strand using the dideoxynucleotide method. The probe used to complete the cloning of the human cDNA library is shown as SEQ ID NO: 37. Finally, a full-length cDNA sequence was assembled which encodes the human PARG. The sequence of the cDNA encoding human PARG is presented in the sequence listing as SEQ ID NO: 3 and the amino acid sequence of human PARG is presented in the sequence listing as SEQ ID NO: 4.

The human PARG sequence shares extensive amino acid sequence homologies with bovine PARG with more than 89% identity. The sequence similarity is also high at the nucleotide level particularly in the region coding for the protein (174ATG-TGA3104). Surprisingly the 5'-untranslated region of the human sequence displays a completely different sequence with an extensive sequence similarity with highly repeated polymorphic DNA sequences found in the human genome such as Alu repetitive elements or variable number of tandem repeats (VNTR).

Example 12

Screening of Mouse Liver 5'-stretch cDNA Libraries

To isolate a PARG cDNA from the mouse liver cDNA library, a probe was designed from the human cDNA clone coding for PARG. Analysis of the bovine and human sequences revealed that PARG was highly conserved between these two species, suggesting that it might also be conserved in the mouse. Based on the restriction map of the human cDNA clone, a region in the human clone was selected, located where the active site of the protein is encoded, that exhibited near identity to its counterpart in the bovine clone. This region, consisting of approximately 800 bases, was excised from the entire human clone by digestion with the restriction endonuclease, HindIII, then purified by agarose gel separation and radiolabeled by random priming.

This probe was used to screen a mouse liver 5'-stretch cDNA library. One clone consistently hybridized with the probe. After two rounds of screening to ensure the purity of the clone the 2.5 kb insert was subcloned into the plasmid pTZ19R and sequenced. Comparison with the sequence of bovine and human PARG showed that this clone had the partial sequence that has extensive similarities to the two other mammalian sequences covering almost entirely the coding region from nucleotide—10 to a few nucleotides from the end of the coding region. A second screen was performed to obtain the missing part of the cDNA using a radioactive probe specifically designed to hybridize with the

region the most 3' of the previous clone to increase the chance to get the missing part of the cDNA.

With this new probe, the same mouse liver cDNA library was screened to obtain a second clone, containing an insert that was about 3 kb. This clone was purified, subcloned and sequenced. The sequence showed that this second clone starts at amino acids 634, extends toward the stop codon to approximately 900 nucleotides into the 3'-non-coding region.

A search of the dBEST database turned up one significant match to a 400 bp fragment cloned from mouse muscularis. This fragment had an exact match to the very tail end of the second clone and exceeded it by 34 bases. This extra extension contained the oligo A sequence as well as the polyadenylation signal. Because there was an exact match, the cDNA sequence was completed using this information coming from the database. The complete cDNA sequence of murine PARG is presented in the Sequence Listing as SEQ ID NO: 5 and SEQ ID NO: 6.

Example 13

Obtaining the Drosophila PARG cDNA

Among the clones obtained from DNA databases searches were several clones from the Drosophila genome sequencing project (European Drosophila Genome Sequencing Consortium) as well as the Drosophila expression sequence TAG sequencing project (67). The EST clone was requested from the University of California Berkeley and obtained. Because the sequence published in the dBEST database was only partial, its sequence was completed in our laboratory and compared to a genomic sequence, part of the distal X chromosome of *Drosophila melanogaster* submitted by Murphy et al., August 1997 which presumably contains the gene of Drosophila PARG. The 768 aa shares less homologies with only 40% identity (48% similarity) mainly located in the catalytic domain of the protein. The domain organization of the protein is also very different with an unknown domain of 20 kDa located Carboxy terminus of the highly conserved active domain. (See FIG. 15). The sequence of the cDNA encoding the Drosophila PARG is presented in the Sequence Listing, as SEQ ID NO: 7 and the amino acid sequence of the Drosophila PARG is presented in the Sequence Listing as SEQ ID NO: 8.

Example 14

Obtaining the *C. elegans* PARG Sequence

This sequence has been obtained by searching the GenBank database with the mammalian PARG protein sequence. A sequence with PARG similarity was found in the cosmid F20C5 (Accession number: Z68161, SEQ ID NO: 38) derived from the *C. elegans* genomic DNA (68). The overall sequence conservation (726aa, MW 83129 Da) with the other PARG sequences is as follows: 32% similarity and 22% identity with the mammalian PARG and 39% similarity and 30% identity with the Drosophila PARG. The sequence is presented in the Sequence Listing as SEQ ID NO: 38 (Genbank accession number CEF20C5). SEQ ID NO: 38 contains 12 exons as follows: exon 1 from 3591 to 3635; exon 2 from 3681 to 4121; exon 3 from 5065 to 5235; exon 4 from 5930 to 6152; exon 5 from 6200 to 6267; exon 6 from 7246 to 7338; exon 7 from 7386 to 7553; exon 8 from 7738 to 7853; exon 9 from 8153 to 8435; exon 10 from 8487 to: 8610; exon 11 from 8662 to 8952; and exon 12 from 9383 to: 9540. The coding sequence of the CePARG protein, which is publicly available from Accession number: Z68161, is referred to in the Sequence Listing as SEQ ID NO: 9. Its corresponding amino acid sequence is referred to in the Sequence Listing as SEQ ID NO: 10. The amino acid sequence of the *C. elegans* PARG is presented on the alignment (FIG. 16)

Example 15

Cloning and Overproduction of the Carboxyl-terminus 69 kDa Domain of Bovine PARG (bPARG) in *E. coli*

As described, above, bovine PARG is encoded by a messenger of 4 kb predicting a protein of 110 kDa, almost twice the size of the purified enzyme (65 kDa). It is also demonstrated that bPARG can be expressed in *E. coli* as an active enzyme either as a 110 kDa or a 65 kDa protein. This result combined with other evidence implies that the active site of PARG is located in the carboxyl-terminal part of the protein. FIG. 11 is a schematic representation of the different clones we have expressed in bacteria. Among them, only the clone designed to express a protein of 69 kDa starting at the amino acid+380 from the sequence of bovine PARG (bPARG_{MNDV}) allowed high level expression as a fusion protein with glutathione-S transferase (GST).

The heterologous expression of bPARG_{MNDV} was conducted as represented in FIG. 12. The 1.8 kb cDNA encoding the 69 kDa carboxyl-terminal part of bovine PARG was amplified by PCR and cloned in the EcoRI site of pGEX-2T vector (Pharmacia) in fusion with GST giving the pGEX-2T-bPARG_{MNDV} plasmid. *E. coli* NM522 cells transformed with the pGEX-2T-bPARG_{MNDV} were induced by addition of IPTG, resulting in expression of a 90 kDa fusion protein. The fusion protein can be conveniently purified using Glutathione-Sepharose and the bPARG_{MNDV} can be released by treatment with thrombin while the GST protein remains bound to the beads of GSH-Sepharose. In this manner milligram amounts of protein can be routinely obtained.

Example 16

Characterization of the Purified 65 kDa Domain and the Generation of Antibodies

The purified bPARG_{MNDV} was characterized by activity gel assays (69) by casting polyacrylamide gels with auto-modified PARP containing [³²P]ADP-ribose polymers. The results demonstrate that the 65 kDa domain expressed in *E. coli* contained enzymatic activity migrating with the same apparent molecular weight as the enzyme purified from bovine thymus. Likewise, a construction expressing bPARG_{MNDV} domain in SF9 insect cells infected with recombinant baculovirus showed activity migrating with the same apparent molecular weight.

The availability of PARG cDNA allows the development of new molecular tools to study this enzyme in its cellular context. Until this work, it was not possible to obtain PARG in sufficient quantities to produce antibodies against the protein. The antibody raised against bovine PARG is able to recognize PARG from other organisms and, thus, will be valuable in characterizing PARG in vivo under defined physiological conditions in many different organisms.

Antibodies against bPARG_{MNDV} overexpressed in *E. coli* were raised in rabbits using the procedure described by Vaitukaitis (70). Specific high affinity antibodies are generated by administration of small doses of immunogens intradermally over a wide anatomic area of the animal. Rabbits were immunized by three injections of 10–50 μg of the Mr 65,000 protein band excised from a preparative SDS polyacrylamide gel. Titer and affinity of sera harvested weekly were followed by conventional methods. Peak affinity was attained in 8 to 10 weeks after primary immunization. For each animal, a preimmune serum was retained as a control.

FIG. 17 shows a Western blot experiment demonstrating the specificity of the resulting PARG anti-serum against the purified bPARG from thymus (lane 1), SF9 protein extract expressing 65 kDa-bPARG_{MNDV} in recombinant baculovirus (lane 2), recombinant 65 kDa-PARG_{MNDV} purified by treatment with thrombin from GSH-Sepharose (lane 3), and an *E.*

coli crude extract expressing the fusion protein GST-65 kDa-PARG_{MNDV} (lane 4). The pre-immune serum did not show reactivity against any of these fractions even at a low dilution (1/250).

Antibodies directed against the 45 kDa-terminal have also been generated using the same strategy used to generate antibodies against the catalytic domain. This involved the overexpression of the 45 kDa protein domain in *E. coli* in a construct designed for easy purification, followed by injection of the purified protein into rabbits. The heterologous expression of PARG45 was conducted by cloning a part (1.1 kb) of the coding region of the cDNA, generated by PCR amplification of the region located between the ATG(267) codon and nucleotide 1400 in the bovine sequence, into the Eco RI site of the bacterial expression vector pGEX-2T (Pharmacia) in fusion with glutathione-S-transferase. *E. coli* NM522 cells transformed with this construct were induced by addition of IPTG, resulting in expression of a 72 kDa fusion protein. The fusion protein was purified using glutathione Sepharose and the PARG45 was released by treatment with thrombin, while the GST protein remained bound to the GSH Sepharose beads. In this manner milligram amounts of protein were obtained. Antibodies against PARG45 overexpressed in *E. coli* were raised in rabbits using the procedure described (71). Specific high affinity antibodies were generated by administration of small doses of immunogens subcutaneously over a wide area of the animal. Rabbits were immunized by three injections of 10–50 μg of the 45 kDa protein band excised from a preparative SDS-polyacrylamide gel. Titer and affinity of sera harvested weekly were followed by conventional methods. Peak affinity was attained in 8 to 10 weeks after primary immunization. For each animal, a preimmune serum was retained as a control.

Example 17

Conservation of PARG in Tissues and Organisms

Tissue and cell extracts from different origins were homogenized in a cold hypotonic lysis buffer containing a cocktail of protease inhibitors and sonicated. SDS and β-mercaptoethanol were added to insure inactivation of any remaining active proteases. Thirty μg of protein from each extract was analyzed by Western-blot using the anti-PARG antibody (FIG. 18). In all of the fractions from bovine tissues, PARG was observed as a major band at 65 kDa. However, less intense, discrete proteins of higher molecular weight were also detected. These proteins may correspond to different forms of PARG; the band of highest molecular weight (about 115 kDa) found in thymus extract likely corresponds to the full-length of PARG (111 kDa) as deduced from the cDNA. Multiple species were detected in cell extracts from mouse fibroblasts, rat PC 12 cells, and SF9 insect cells. This result shows that the sequence of PARG is well conserved phylogenetically. Moreover, the conservation includes multiple molecular forms of the protein.

Example 18

Regulation of the Expression of PARG

In the metabolism of ADP-ribose polymers, the activities of PARP and PARG are closely related. Soon after polymer has been synthesized by PARP following DNA damage, it is extensively degraded by PARG. The net result is that the polymer has a very short half life. The close relationship between the two proteins suggests a possible mode of regulation in which PARG expression depends on the presence of PARP. In order to test if the presence or the absence of PARP influences the expression of PARG, a Western Blot

experiment was performed with cell extracts from mouse fibroblasts of different PARP genotypes (72).

Cell extract (30 μ g) from mouse cells with PARP +/+, PARP +/- and PARP -/- genotypes were separated by SDS-PAGE, transferred to a membrane and probed with the antibodies indicated. The results are shown in FIG. 19. In FIG. 19, purified PARG (50 ng) from bovine thymus (lane 1), 30 μ g of protein of a total extract from PARG recombinant baculovirus infected SF9 cells (lane 2), 150 ng of purified recombinant PARG produced in the bacteria (lane 3) and 30 μ g of protein of a crude extract from *E. coli* NM522 transformed with pGEX-2T-bPARG_{MNDV} 2 h after induction by IPTG (lane 4) were separated on a 0.1% SDS-12% polyacrylamide gel, then transferred on nitrocellulose, and incubated with a 1/5000 dilution of the rabbit polyclonal antiserum raised against the 65 kDa domain of bPARG. Proteins were revealed by immunofluorescence with the ECL detection kit (Amersham) and autoradiography. Panel A is a western blot of PARP in cells of varying PARP genotype showing the results of the analysis using anti-PARP antibodies. The amount of PARP expressed varies as expected dependant upon the genotype of the cell line with the PARP -/- cell line producing no detectable amount of PARP. Panel B is western blot of PARG from various tissues using an anti-PARG antibody. It shows that the level of PARP is variable. The amount of PARG present in the cell extracts was not dependent upon the PARG genotype of the cell. Further support for this view is provided by the results of the PARG activity assay presented in panel C. The specific activity of PARG detected in the extracts showed no significant difference among the three genotypes.

Example 19

Preparation PARG Gene Ablation (Knockout) Animals

One embodiment of the present invention is experimental animals with targeted mutations in the PARG gene. These animals may be constructed using standard techniques and the cDNA sequence of the PARG. In the following example, a mouse containing a targeted mutation in the PARG gene is constructed. Those skilled in the art will readily appreciate that other experimental animals, including but not limited to rats, guinea pigs, hamsters and the like, may be constructed using similar techniques. The construction of animals with disrupted genes may be accomplished using standard techniques such as those described by Moreadith (73). Further, cells lines, construction kits, and protocols for knockout mice are available from commercial suppliers such as Stratagene (Stratagene 1999 catalog, La Jolla, Calif.). Commercial services such as Lexicon Genetics (The Woodlands, Tex.) and Chrysalis DNX Transgenic Sciences (Princeton, N.J.) also offer complete ES cell knockout mice production services.

A genomic clone of the murine PARG enzyme may be isolated from a genomic library by screening with a probe derived from the cDNA sequence of PARG. A mouse 129/SV genomic library (Stratagene) containing mouse genomic sequences in λ phage was screened using a 2.49 kb fragment of the mouse PARG cDNA as a probe. A partial restriction map of one positive clone thus isolated, R1, is provided in FIG. 20. The R1 clone contains the genomic sequence corresponding to the 5'-most end of the murine cDNA. The clone was subcloned into pBluescript as three fragments. The plasmid containing the 5'-end contained a 2.8 kb NotI-EcoRI fragment and was designated p2.8R. The fragment containing the central portion contained a 3.5 kb EcoRI fragment and was designated p3.5R. The plasmid containing the 3'-end of the gene contained a 7.0 kb EcoRI-

NotI fragment and was designated 7.0R. Sequencing the resulting plasmids revealed that p2.9R contained no sequences corresponding to the cDNA, p3.5R contains a 1.5 kb promoter region and/or untranslated region and exon 1 coding for 72 amino acids including the initiation ATG codon and p7.0R contains at least 4 additional exons. Gene targeting vectors may be constructed using both p3.5R and p7.0R.

A gene targeting vector may contain one or more selection genes flanked by genomic sequences. The targeting vector is introduced into the genome by homologous recombination resulting in the incorporation of the selection gene into the genome of the cell. The mouse PARG gene was targeted using a "conditional" inactivation procedure outlined in FIG. 21. This approach allows the production of viable animals even if the disrupted gene results in a lethal phenotype since the gene is not disrupted until a second "conditional" recombination event is induced.

A lox-P sequence may be inserted into the first intron. A cassette expressing the neomycin resistance gene (neo) and the thymidine kinase gene (TK) flanked by two additional lox-P sites may be placed in intron 2. In the presence of Cre recombinase, recombination will occur between two lox-P sites thereby deleting the genomic sequences present between the sites. A MCI-DTA cassette is ligated at the 3'-end of the vector to reduce random integration of the vector into the genome.

The targeting vector may be introduced into embryonic stem cells by any method known to those skilled in the art such as transfection, lipofection or electroporation. In a preferred embodiment, the targeting vector will be introduced into embryonic stem cells by electroporation. After homologous recombination, cells containing the neo gene will be selected for using G418. Selected cells will then be analyzed by PCR and Southern blot.

To generate mutant alleles of PARG, the positive embryonic stem cell clones identified will be transfected with a plasmid expressing Cre recombinase. The action of Cre recombinase will result in three different mutant alleles. Mutant allele I contains a deletion in exon 2 but still maintains the selection genes neo and TK. Mutant allele II contains the genomic sequence for exon 2 flanked by two lox-P sites (exon 2 is said to be "floxed") and does not contain the selection genes. Mutant allele III has a deletion of the genomic sequences and does not contain the selection genes.

Mice containing each of the three mutant alleles may be constructed by microinjecting embryonic stem cells containing the mutant allele into blastocytes resulting in the production of chimeric and mutant mice. Mice homozygous in mutant allele I or III will be null mutants in that they will be unable to express a functional PARG enzyme due to the loss of required genomic sequences. In the absence of Cre recombinase, mice containing mutant allele II will express a wild type protein. In the presence of the Cre recombinase, the PARG will lose exon 2, thus producing an inactive protein. To inactivate the gene, these mice will be bred to mice expressing Cre recombinase under the control of a tissue specific promoter. This will result in mice expressing PARG in some tissues and not expressing PARG in others. Mice homozygous in mutant allele II, will be valuable for evaluating the role of PARG in specific tissues.

Although the present invention has been described with reference to certain examples for purposes of clarification and illustration. It should be appreciated that certain obvious improvements and modifications can be practiced within the scope of the appended claims and their equivalents. Other

embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein.

All U.S. Patents GenBank sequence listings, and other references noted herein for whatever reason are specifically incorporated by reference. The specification and examples should be considered exemplary only with the true scope and spirit of the invention indicated by the following claims.

1. Miwa, M. et al. (1971) *J. Biol. Chem.* 246, 6362–6364; Ueda, K. et al. (1972) *Biochem. Biophys. Res. Commun.* 46, 516–523
2. Oka, J. et al. (1984) *J. Biol. Chem.* 259, 986–995.
3. Gaal, J. et al. (1987) *Trends in Biol. Sci.*, 12, 129.
4. Nudka, N. et al. (1980) *Eur. J. Biochem.* 105, 525–530; Jacobson, E. et al. (1985) *Carcinogenesis* 6, 715–718; Kupper, J. et al. (1990) *J. Biol. Chem.* 265, 18721–18724; Ding, R. et al. (1992) *J. Biol. Chem.* 267, 12804–12812.
5. Jacobson, E. et al. (1985) in *ADP-ribosylation of Proteins* (Althaus, F. R., Hiltz, H., and Shall, S., eds) pp. 277–283, Springer-Verlag, Berlin; Lubet, R. et al. (1986) *Carcinogenesis* 7, 71–75; Kasid, U. et al. (1986) *Carcinogenesis* 7, 327–330.
6. Berger, N. (1985) *Radiat. Res.* 101, 4–15.
7. Kaufmann, S. et al. (1993) *Cancer Res.* 53, 3976–3985; Lazebnik, Y. et al. (1994) *Nature* 371, 346–347.
8. de Murcia, G. et al. (1994) *Trends Biochem. Sci.* 19, 172–176.
9. BioWorld Today, Apr. 29, 1994, p. 3.
10. Thomassin, H. et al. (1990) *Nucleic Acids Res.* 18, 4691–4694.
11. Karlin, Samuel and Stephen F. Altschul (1993). *Proc. Natl. Acad. Sci. USA* 90:5873–7
12. Henikoff and Henikoff *Proc. Natl. Acad. Sci. USA* 89:10915–19, 1992
13. Altschul, Stephen F. (1991). *J. Mol. Biol.* 219:555–65.
14. Cohen, J. S., 1989; Weintraub, H. M., 1990
15. N. Sarver et al., 1990
16. Orlandi, R. et al., (1989) *Proc. Natl. Acad. Sci.* 86, 3833–3837; Winter, G et al., (1991) *Nature* 349, 293–299
17. R. Sobol and K. Scanlon eds., available at www.appleton-lange.com.
18. e.g., at least, but not limited to U.S. Pat. Nos. 5,797,870, 5,804,383, 5,670,161, 5,645,829, 5,741,486, 5,836,905, 5,843,069, 5,827,216, 5,871,464, 5,702,384, 5,810,888, 5,787,900, 5,752,515, 5,674,192, 5,658,955, 5,656,465, 5,547,932, 5,873,904 5,792,651, 5,772,888, 5,641,750, 5,641,749, and 5,626,561
19. Kaltenbock, B. et al. (1998) *Biotechniques*, 24, 202–206.
20. Moran, P. et al. (1998) *Biotechniques*, 24, 206–212.
21. Orita, M. et al. (1989) *PNAS USA*, 86, 2766–2770.
22. Liu, Q. et al. (1998) *Biotechniques*, 24, 140–147.
23. *Proceeding from the Sixth International Symposium on Human Identification* 1995 (ISBN 1-882274-55-5).
24. Kohler and Milstein, *Nature*, 256:495–497 (1975).
25. Geysen, et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:3998–4002.
26. Ménard, et al. (1987) *Biochem. Cell Biol.* 65, 668–673.
27. Hatakeyama, K. et al. (1986) *J. Biol. Chem.* 261, 14902–14911.
28. Thomassin, H. et al. (1990) *Nucleic Acids Res.* 18, 4691–4694.
29. Ménard, L. et al. (1987) *Biochem. Cell Biol.* 65, 668–673.
30. Bradford, M. (1976) *Anal. Biochem.* 72, 248–254.
31. Laemmli, U. (1970) *Nature* 227, 680–685.
32. Althaus, F. et al. (1987) *Molecular Biology, Biochemistry and Biophysics*, Vol. 37, Springer-Verlag, Berlin.

33. Mead, D. et al. (1986) *Protein Eng.* 1, 67–74.
34. Shaw, et al. (1986) *Cell* 46, 659–667.
35. Kozak (1987) *Nucleic Acids Res.* 15, 8125–8148.
36. Wilson, et al. (1994) *Nature (London)* 368, 32–38.
37. Brendel, et al. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2002–2006.
38. Robbins, et al. (1991) *Cell* 64, 615–623.
39. Schreiber, et al. (1992) *EMBO J.* 11, 3263–3269.
40. Uchida, K. et al. (1987) *Biochem. Biophys. Res. Commun.* 148, 617–622; Cherney, B. et al. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 8370–8374; Kurosaki, T. et al. (1987) *J. Biol. Chem.* 262, 15990–15997.
41. Huppi, K. et al. (1989) *Nucleic Acids Res.* 17, 3387–3401.
42. Saito, I. et al. (1990) *Gene (Amst.)* 90, 249–254.
43. Ittel, M.-E. et al. (1991) *Gene (Amst.)* 102, 157–164.
44. Saulier-Le Drean, B. (1992) *Poly(ADP-ribose) Polymerase in Xenopus lae-vis*, Ph.D. thesis, Université De Rennes, France.
45. Uchida, K. et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 3481–3485.
46. Masutani, M. et al. (1994) *Eur. J. Biochem.* 220, 607–614.
47. Schreiber, et al. (1992) *EMBO J.* 11, 3263–3269.
48. Lazebnik, et al. (1994) *Nature (London)* 371, 346–347.
49. Brochu, et al. (1994) *Biochem. Biophys. Acta* 1219, 342–350.
50. Tanuma, et al. (1986) *J. Biol. Chem.* 261, 965–969; Uchida, et al. (1993) *J. Biol. Chem.* 268, 3194–3200.
51. Maruta, et al. (1991) *Biochemistry* 30, 5907–5912.
52. Ménard, L. et al. (1987) *Biochem. Cell Biol.* 65, 668–673.
53. Ménard, L. et al. (1987) *Biochem. Cell Biol.* 65, 668–673.
54. Slama, J. et al. (1995) *J. Med. Chem.* 38, 389–393; Slama, J. et al. (1995) *J. Med. Chem.* 38, 4332–4336.
55. Slama, J. et al. (1995) *J. Med. Chem.* 38, 389–393; Slama, J. et al. (1995) *J. Med. Chem.* 38, 4332–4336.
56. Ménard, L. et al. (1987) *Biochem. Cell Biol.* 65, 668–673.
57. Althaus, F. et al. (1987) *Molecular Biology, Biochemistry and Biophysics*, Vol. 37, Springer-Verlag, Berlin.
58. Brochu, G. et al. (1994) *Anal. Biochem.* 218, 265–272.
59. Brochu, G. et al. (1994) *Anal. Biochem.* 218, 265–272.
60. Moreadith, et al. (1997) *J. Mol. Med.* 75, 208–216.
61. Tanuma, et al. (1986) *J. Biol. Chem.* 261, 965–969 and Uchida, et al. (1993) *J. Biol. Chem.* 268, 3194–3200.
62. Hatakeyama, et al. (1986) *J. Biol. Chem.* 261, 14902–14911; Thomassin, et al. (1990) *Nucleic Acids Res.* 18, 4691–4694; and Maruta, et al. (1991) *Biochemistry* 30, 5907–5912.
63. Sambrook, et al. (1992) *MOLECULAR CLONING: A LABORATORY MANUAL*, Cold Spring Harbor Laboratory Cold Spring harbor, N.Y.
64. Feinberg, et al. (1983) *Anal. Biochem.* 132, 6–13.
65. Boguski, 1995
66. Sambrook, et al. (1992) *MOLECULAR CLONING: A LABORATORY MANUAL*, Cold Spring Harbor Laboratory Cold Spring Harbor, N.Y.
67. BDGP/HHMI Drosophila EST Project, University of California Berkeley, EST@fruitfly.berkeley.edu
68. Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB101RQ, England and Department of Genetics, Washington University, St. Louis, Mo. 63110, USA. E-mail: jes@sanger.ac.uk
69. Brochu G. et al. (1994) *Anal. Biochem.* 218, 265–272.

70. Vaitukaitis (1981) *Methods in Enzymology* 73, 46–52.
71. Vaitukaitis (1981) *Methods in Enzymology* 73, 46–52.

72. Wang, et al. (1995) *Genes & Dev.* 9, 509–520.
73. Moreadith, et al. (1997) *J. Mol. Med.* 75, 208–216.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 38

<210> SEQ ID NO 1

<211> LENGTH: 4070

<212> TYPE: DNA

<213> ORGANISM: *Bos taurus*

<220> FEATURE:

<400> SEQUENCE: 1

```

accgaaagt gaacgaagcc cgaatcagaa cggctcatcc tgaggctggt a gggtgccgg      60
tggaagaggg aaggcaggcg tctggatagg gcctggttcg ggaggctgtc a gagcaggag      120
ctgcagaagc agtcagcggc agagggggca tggtgccggg aggcaccgag g agggggcgc      180
agtccgtccc tcccagggtt agtgaatgag gctctacgcc cgggctggcc c ggagactca      240
gtgctgcggg tcccagcatg agtgccggcc ccggctgtga gccctgcacc a agcgacccc      300
gctgggagcgc cgctgcaact tctccgcggc ccgcctcgga cggccggagc t tccccggca      360
ggcagaggcg cgtcctcgat tccaaggacg ctccggtgca gttcagggtc c cgccgtcct      420
cgtcaggctg cggcctgggc cggggcggac agcaccgagg cagcggcacc t ctcttgttt      480
tcaaacagaa gactataacc agttggatgg aactaaagg aatcaagaca g ttgaatcag      540
aaagtttgca tagtaagaa aacaacaata caagagaaga atccatgatg a gttctgtac      600
aaaaagataa cttttatcaa cataacatgg aaaaattaga aaatgtttct c agctaggtt      660
ttgataagtc accagttgaa aaaggtacac agtatttgaa gcagcatcag a ctgcccgtta      720
tgtgtaagtg gcagaatgaa gggccacact cagaacggct tttgaaagt g aacctccag      780
cggtaactct ggtaccagag cagttcagta atgctaattg cgatcagtcg t ccccaaagg      840
atgatcacag tgacacaaat agtgaggaga gtagagataa tcagcagttt t tgacacatg      900
taaaagcttc gaatgcaaag cagacgatgg aagatgaaca gggcagagaa g ccagaagcc      960
accagaagtg tggaaggtc tgccatcctg cagaagcctg tgcagggtgt c agcaggagg      1020
agacagacgt ggtgtccgag agccccttgt cggacactgg ctctgaggat g ttggtactg      1080
gactgaaaaa tgccaacaga ttgaatagac aagaaagtag tctaggaaat t ctcctccat      1140
ttgagaaaga aagtgaaact gagtcaccaa tggatgtaga taattccaaa a atagttgtc      1200
aggattcaga agcagatgaa gagacaagtc caggttttga tgaacaggaa g atagcagtt      1260
ctgctcaaac agcaataaaa ccttcaaggt tccaaccaag agaagctgac a ctgagttga      1320
ggaagcggtc ctctgctaag ggaggtgaga ttcgattaca tttccaattt g aaggaggag      1380
agagtcgagc tggaatgaat gatgtgaatg ccaaacgacc tggaagtact t ctagcctga      1440
atgtagagtg cagaaattct aagcaacatg ggagaaagga ttctaaaatc a cagatcatt      1500
tcatgagagt gcccaagaca gaggacaaaa gaaaagaaca atgtgaaatg a aacatcaaa      1560
gaacagaaag gaagatccct aaatacattc cacctcacct tctccagat a agaaatggc      1620
ttggaactcc tattgaggag atgaggagaa tgccaaggtg tgggatccgg c tgcctccct      1680
tgagaccatc tgccaatcac acagtgacta ttcgggtaga tcttttcgca a taggagaag      1740
ttcctaaacc tttccaaca cattttaaag atttgtggga caacaagcat g ttaagatgc      1800
cttgttcaga aaaaacttg tacctctgtg aagatgagaa tggtgagcga g ctgcaggca      1860

```

-continued

gccggtggga	actcattcag	actgcacttc	tcaacaggct	cactcggccc	c	agaacctga	1920
aggatgctat	tctgaagtac	aatgtggcat	attctaagaa	atgggacttt	a	cagctttga	1980
ttgatttctg	ggataaggta	ctagaagaag	cagaagctca	acacttgat	c	agtccatct	2040
tgctgatata	ggtgaaaatt	gcactctgtc	tgccaaatat	ttgtaccag	c	caataccac	2100
tcttgaaaca	gaagatgaat	cattccatca	caatgtcaca	ggaacagatt	g	ccagtcttt	2160
tagctaattc	tttcttctgc	acgtttccac	gacgcaatgc	caagatgaaa	t	cagagtatt	2220
ccagttatcc	agatattaac	ttcaatcggt	tgtttgaagg	acgttcatca	a	ggaaccag	2280
agaagcttaa	aacgctcttc	tgctacttta	gaagagtcac	agagaaaaaa	c	ccactgggt	2340
tggtgacatt	cacaagacag	agtcttgaag	atcttccaga	gtgggaaaga	t	gtgaaaaac	2400
tcttgactcg	actgcatgtc	acttacgaag	gtaccataga	aggaaacggc	c	agggcatgc	2460
tacaggtgga	ttttgcaaac	cgtttcgttg	gaggtggtgt	aaccagtgc	a	gacttggtc	2520
aagaagaaat	ccgtttttta	atcaacctcg	agttgattgt	ttcacggctc	t	tactgagg	2580
tgctggatca	caatgaatgt	cttatcatca	caggtactga	gcagtacagt	g	aatcacag	2640
gctatgccga	aacataccgc	tggtcccgga	gccatgaaga	caggagcga	a	gggacgact	2700
ggcagaggcg	cacgactgag	atcgtcgcca	tcgacgcctc	ccacttcaga	c	gctacctcg	2760
accagtttgt	gcccagaaag	atcagacggg	agcttaacaa	ggcttactgt	g	gatttcttc	2820
gtcctggagt	ttcttcagag	aacctgtctg	cagtggttac	aggaaactgg	g	gctgtggtg	2880
cctttggggg	tgatgctaga	ctaaaagcct	taatacagat	cctggcagct	g	ctgtagctg	2940
agcgagacgt	ggtttatttc	acctttgggg	actcagaact	gatgagagac	a	tttacagca	3000
tgcatatact	cctcactgag	aggaaactga	ctgttgaga	agtatataag	c	tgctgctac	3060
gatattacaa	tgaagaatgc	agaaactgct	ccacccccgg	accagacatc	a	agctttatc	3120
cattcatata	ccatgcagtt	gagtcctgta	cacagaccac	caaccagccg	g	gacaaagga	3180
cggtggcctg	aggagccaag	tgactagacg	ctccccactt	gtgtaacaag	a	aggtgtgac	3240
gtgtgaaact	acatgatata	catgtgtata	taatccgctg	ttgtaggcaa	g	gatgcagtc	3300
ccttccgccc	atgcagctgt	cagtacatct	gogcctcctc	catcccgact	t	acatagact	3360
gagacatact	ttgtttcttt	ttttttctat	ttcagccctg	attcttttat	t	tttctttct	3420
tttgcccata	agacttcttg	tgaaatttca	tcagagtttg	tgctcagcct	g	gcaggtgtc	3480
ttttttgatg	cctaaatata	caaatcacct	ctgcagctag	cagatgccac	g	gaaggtggt	3540
ggaaccctag	gagctgtaac	tgagtctgct	gcagatctcc	ctctgagcct	c	tcaccctca	3600
ccctattata	attgtggtgg	tgaggttttt	ttgatttttg	aaataagagt	t	gggtttggt	3660
aaataatata	gatctcctag	gtaagagttt	ttatatttaa	gaatactttt	c	aaaaagtta	3720
ttttgagata	tcacctttat	ttgtaatggt	aatttgctg	tcccttttcc	c	ctgatcaat	3780
ttgtattgac	tgtttttgg	aattgaccac	aatgaaagga	aatatgagaa	t	aagagtttc	3840
ccaaatggtg	tttaaaaaca	aacaggttca	agacacgcga	aggacctcgt	t	tcctgggat	3900
tttttttctt	tttctttttt	tgaattagga	ttattgtttg	ttccttggtg	c	ttgagacat	3960
attcatataa	ccaaagttta	ggaactggga	acttcgtggt	gatttgtaca	t	attgaagtt	4020
tctctgtgac	tcaaagggtta	tgtagttaat	aaattttcat	taacaaaaaa			4070

<210> SEQ ID NO 2

<211> LENGTH: 977

<212> TYPE: PRT

<213> ORGANISM: Bos taurus

-continued

<220> FEATURE:

<400> SEQUENCE: 2

Met Ser Ala Gly Pro Gly Cys Glu Pro Cys Thr Lys Arg Pro Arg Trp
1 5 10 15
Asp Ala Ala Ala Thr Ser Pro Pro Ala Ser Asp Ala Arg Ser Phe
20 25 30
Pro Gly Arg Gln Arg Arg Val Leu Asp Ser Lys Asp Ala Pro Val Gln
35 40 45
Phe Arg Val Pro Pro Ser Ser Gly Cys Ala Leu Gly Arg Ala Gly
50 55 60
Gln His Arg Gly Ser Ala Thr Ser Leu Val Phe Lys Gln Lys Thr Ile
65 70 75 80
Thr Ser Trp Met Asp Thr Lys Gly Ile Lys Thr Val Glu Ser Glu Ser
85 90 95
Leu His Ser Lys Glu Asn Asn Asn Thr Arg Glu Glu Ser Met Met Ser
100 105 110
Ser Val Gln Lys Asp Asn Phe Tyr Gln His Asn Met Glu Lys Leu Glu
115 120 125
Asn Val Ser Gln Leu Gly Phe Asp Lys Ser Pro Val Glu Lys Gly Thr
130 135 140
Gln Tyr Leu Lys Gln His Gln Thr Ala Ala Met Cys Lys Trp Gln Asn
145 150 155 160
Glu Gly Pro His Ser Glu Arg Leu Leu Glu Ser Glu Pro Pro Ala Val
165 170 175
Thr Leu Val Pro Glu Gln Phe Ser Asn Ala Asn Val Asp Gln Ser Ser
180 185 190
Pro Lys Asp Asp His Ser Asp Thr Asn Ser Glu Glu Ser Arg Asp Asn
195 200 205
Gln Gln Phe Leu Thr His Val Lys Leu Ala Asn Ala Lys Gln Thr Met
210 215 220
Glu Asp Glu Gln Gly Arg Glu Ala Arg Ser His Gln Lys Cys Gly Lys
225 230 235 240
Ala Cys His Pro Ala Glu Ala Cys Ala Gly Cys Gln Gln Glu Glu Thr
245 250 255
Asp Val Val Ser Glu Ser Pro Leu Ser Asp Thr Gly Ser Glu Asp Val
260 265 270
Gly Thr Gly Leu Lys Asn Ala Asn Arg Leu Asn Arg Gln Glu Ser Ser
275 280 285
Leu Gly Asn Ser Pro Pro Phe Glu Lys Glu Ser Glu Pro Glu Ser Pro
290 295 300
Met Asp Val Asp Asn Ser Lys Asn Ser Cys Gln Asp Ser Glu Ala Asp
305 310 315 320
Glu Glu Thr Ser Pro Gly Phe Asp Glu Gln Glu Asp Ser Ser Ser Ala
325 330 335
Gln Thr Ala Asn Lys Pro Ser Arg Phe Gln Pro Arg Glu Ala Asp Thr
340 345 350
Glu Leu Arg Lys Arg Ser Ser Ala Lys Gly Gly Glu Ile Arg Leu His
355 360 365
Phe Gln Phe Glu Gly Gly Glu Ser Arg Ala Gly Met Asn Asp Val Asn
370 375 380
Ala Lys Arg Pro Gly Ser Thr Ser Ser Leu Asn Val Glu Cys Arg Asn
385 390 395 400

-continued

Ser Lys Gln His Gly Arg Lys Asp Ser Lys I le Thr Asp His Phe Met
 405 410 415

Arg Val Pro Lys Ala Glu Asp Lys Arg Lys G lu Gln Cys Glu Met Lys
 420 425 430

His Gln Arg Thr Glu Arg Lys Ile Pro Lys T yr Ile Pro Pro His Leu
 435 440 445

Ser Pro Asp Lys Lys Trp Leu Gly Thr Pro I le Glu Glu Met Arg Arg
 450 455 460

Met Pro Arg Cys Gly Ile Arg Leu Pro Pro L eu Arg Pro Ser Ala Asn
 465 470 475 480

His Thr Val Thr Ile Arg Val Asp Leu Leu A rg Ile Gly Glu Val Pro
 485 490 495

Lys Pro Phe Pro Thr His Phe Lys Asp Leu T rp Asp Asn Lys His Val
 500 505 510

Lys Met Pro Cys Ser Glu Gln Asn Leu Tyr P ro Val Glu Asp Glu Asn
 515 520 525

Gly Glu Arg Ala Ala Gly Ser Arg Trp Glu L eu Ile Gln Thr Ala Leu
 530 535 540

Leu Asn Arg Leu Thr Arg Pro Gln Asn Leu L ys Asp Ala Ile Leu Lys
 545 550 555 560

Tyr Asn Val Ala Tyr Ser Lys Lys Trp Asp P he Thr Ala Leu Ile Asp
 565 570 575

Phe Trp Asp Lys Val Leu Glu Glu Ala Glu A la Gln His Leu Tyr Gln
 580 585 590

Ser Ile Leu Pro Asp Met Val Lys Ile Ala L eu Cys Leu Pro Asn Ile
 595 600 605

Cys Thr Gln Pro Ile Pro Leu Leu Lys Gln L ys Met Asn His Ser Ile
 610 615 620

Thr Met Ser Gln Glu Gln Ile Ala Ser Leu L eu Ala Asn Ala Phe Phe
 625 630 635 640

Cys Thr Phe Pro Arg Arg Asn Ala Lys Met L ys Ser Glu Tyr Ser Ser
 645 650 655

Tyr Pro Asp Ile Asn Phe Asn Arg Leu Phe G lu Gly Arg Ser Ser Arg
 660 665 670

Lys Pro Glu Lys Leu Lys Thr Leu Phe Cys T yr Phe Arg Arg Val Thr
 675 680 685

Glu Lys Lys Pro Thr Gly Leu Val Thr Phe T hr Arg Gln Ser Leu Glu
 690 695 700

Asp Phe Pro Glu Trp Glu Arg Cys Glu Lys L eu Leu Thr Arg Leu His
 705 710 715 720

Val Thr Tyr Glu Gly Thr Ile Glu Gly Asn G ly Gln Gly Met Leu Gln
 725 730 735

Val Asp Phe Ala Asn Arg Phe Val Gly Gly G ly Val Thr Ser Ala Gly
 740 745 750

Leu Val Gln Glu Glu Ile Arg Phe Leu Ile A sn Pro Glu Leu Ile Val
 755 760 765

Ser Arg Leu Phe Thr Glu Val Leu Asp His A sn Glu Cys Leu Ile Ile
 770 775 780

Thr Gly Thr Glu Gln Tyr Ser Glu Tyr Thr G ly Tyr Ala Glu Thr Tyr
 785 790 795 800

Arg Trp Ala Arg Ser His Glu Asp Arg Ser G lu Arg Asp Asp Trp Gln
 805 810 815

Arg Arg Thr Thr Glu Ile Val Ala Ile Asp A la Leu His Phe Arg Arg

-continued

	820		825		830	
Tyr	Leu Asp	Gln Phe	Val	Pro Glu	Lys Ile A	rg Arg Glu
	835			840		845
Ala	Tyr Cys	Gly Phe	Leu Arg	Pro Gly	Val S er	Ser Glu
	850		855			860
Ala	Val Ala	Thr Gly	Asn Trp	Gly Cys	Gly A la	Phe Gly
865			870		875	880
Arg	Leu Lys	Ala Leu	Ile Gln	Ile Leu	Ala A la	Ala Val
		885			890	895
Asp	Val Val	Tyr Phe	Thr Phe	Gly Asp	Ser G lu	Leu Met
		900		905		910
Tyr	Ser Met	His Thr	Phe Leu	Thr Glu	Arg L ys	Leu Thr
	915			920		925
Val	Tyr Lys	Leu Leu	Leu Arg	Tyr Tyr	Asn G lu	Glu Cys
	930		935			940
Ser	Thr Pro	Gly Pro	Asp Ile	Lys Leu	Tyr P ro	Phe Ile
945			950		955	960
Val	Glu Ser	Cys Thr	Gln Thr	Thr Asn	Gln P ro	Gly Gln
		965		970		975

Ala

<210> SEQ ID NO 3
 <211> LENGTH: 4069
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:

<400> SEQUENCE: 3

```

ggcgtctggg aagtgaggag cgtctctgcc tggcagaggc tgcaatctct g cactttggg      60
gggccaagcg aggcgctgag aaggacgcgc agtccatctc tctcaggtta g tgaatgag      120
gctctccgcy gggccggccc ggggacatgt cgctgctggt cccagcatga a tgcgggccc      180
cggctgtgaa ccctgcacca aagcgaccog ctggggcgcc gctacaactt c gccggctgc      240
ttcggacgcc cggagctttc cgagcaggca gaggcgcgtc ctcgacccca a ggacgtca      300
cgtgcagttc aggttccacc cgtcctcgcc agcctgcgtc ccagggcagg c gggacagca      360
cagaggcagc gccacctcgc ttgttttcaa acaaaagact attaccagtt g gatggacac      420
taaaggaatc aagacagcgg aatcagaaa tttggatagt aaagaaaaca a caatacaag      480
aatagaatcc atgatgagtt ctgtacaaaa agataacttt taccaacata a tgtagaaaa      540
attagtaaat gtttctcagc taagtcttga taagtcactc actgaaaaaa g tacacagta      600
tttgaaccag catcagactg cagcaatgtg taagtggcaa aatgaaggga a acacacgga      660
gcagcttttg gaaagtgaac ctcaaacagt aaccttgta ccagagcagt t tagtaatgc      720
taacattgat cggtcacctc aaaatgatga tcacagtgc acagatagtg a agagaatag      780
agacaatcaa cagtttctca caactgtaaa gcttgcaaat gcaaagcaga c tacggaaga      840
tgaacacgcc agagaagcca aaagccacca gaagtgcagc aagtcttgcc a tcctgggga      900
agactgtgca agttgtcagc aagatgagat agacgtggtg ccaaagagtc c attgtcaga      960
tgttggctct gaggatgttg gtactgggtc aaaaaatgac acaaaattga t tagacaaga     1020
aagttgccta ggaaattctc ctocatttga gaaggaaagt gaaccggaat c accgatgga     1080
tgtggataat tctaaaaata gttgtcaaga ctcagaagca gatgaggaga c aagtccagg     1140
ttttgatgaa caagaagatg gtagttcctc ccaaacagca aataaacctt c aaggttcca     1200
    
```

-continued

agcaagagac gctgacattg aatntagaa acggtactct actaaggcg g tgaagttag	1260
attacatttc caatttgaag gaggagagag tcgactgga atgaatgatt t aaatgctaa	1320
actacctgga aatatttcta gcctgaatgt agaatgcaga aattctaagc a acatgaaa	1380
aaaggattct aaaatcacag atcatttgat gagactgcc aaagcagag a cagaagaaa	1440
agaacagtgg gaaaccaaac atcaaagaac agaaaggaag atccctaaat a cgttccacc	1500
tcacctttct ccagataaga agtggcttgg aactcccatt gaggagatga g aagaatgcc	1560
tcggtgtggg atccggctgc ctctcttgag accatctgcc aatcacacag t aactattcg	1620
ggtagatctt ttgcgagcag gagaagtcc taaacctttt ccaacacatt a taaagattt	1680
gtgggataac aagcatgtta aatgccttg ttcagaacaa aatttgtacc c agtggaga	1740
tgagaatggt gagcgaactg cggggagcgg gtgggagctc attcagactg c acttctcaa	1800
caaatttaca cgaccccaa acttgaagga tgctattctg aaatacaatg t ggcatattc	1860
taagaaatgg gactttacag ctttgatcga tttctgggat aaggacttg a agaagcaga	1920
agctcaacat ttatatcagt ccatcttgcc tgatatggtg aaaattgcac t ctgtctgcc	1980
aaatatttgc acccagccaa taccactcct gaaacagaag atgaatcatt c catcacaat	2040
gtcgcaggaa cagattgcca gtcttttagc taatgcttcc tctgcacat t tccacgacg	2100
aaatgctaag atgaaatcgg agtattctag ttaccagac attaaactca a tcgattgtt	2160
tgagggacgt tcatcaagga aaccggagaa acttaaaacg ctcttctgct a ctttagaag	2220
agtcacagag aaaaaaccta ctgggttggg gacatttaca agacagagtc t tgaagattt	2280
tccagaatgg gaaagatgtg aaaaaccctt gacacgattg catgtcactt a cgaaggtag	2340
catagaagaa aatggccaag gcatgctaca ggtggatttt gcaaatcgtt t tgttgagg	2400
tggtgtaacc agtgcaggac ttgtgcaaga agaaatccgc tttttaatca a tcctgagtt	2460
gattatttca cggctcttca ctgaggtgct ggatcacaat gaatgtctaa t tatcacagg	2520
tactgagcag tacagtgaat acacaggcta tgctgagaca tatogttggt c ccggagcca	2580
cgaagatggg agtgaaaggc acgactgcga gcgagcgtgc actgagatcg t tgccatcga	2640
tgctcttcaac ttcagacgct acctcgatca gtttgtgctt gagaaaatga g acgagcgt	2700
gaacaaggct tactgtggat ttctcctgcc tggagtttct tcagagaatc t ttctgcagt	2760
ggccacagga aactggggct gtgggtgcctt tgggggtgat gccaggttaa a agccttaat	2820
acagatattg gcagctgctg cagctgagcg agatgtgggt tatttcacct t tggggactc	2880
agaattgatg agagacattt acagcatgca cttttcctt actgaaagga a actcactgt	2940
tgagatgtg tataagctgt tgctacgata ctacaatgaa gaatgcagaa a ctgttccac	3000
ccctggacca gacatcaagc tttatccatt catataccat gctgtcgagt c ctgtgcaga	3060
gaccgctgac cattcagggc aaaggacagc gacctgagga gccgagcga t agcatctcc	3120
tcccacctcc caccagagac gtctgtttg agctgtcagg tgtaatatat g aattgactt	3180
aagttaatat aaatgtgtac ataatccaca tttgtagtca aggacgcaat c tcttccaca	3240
catgtgcagt tgctcagttg tacatctaaa ctccctccat cctgactcac g tggacttag	3300
atatgtttt tttctatttt ctctatttc agtttttcat tctttgatgt t tatttcttt	3360
tgtccatcag atctcttgtg aaatcccatg gaaggttgtg ctgagctgtc g ggtctcttt	3420
cttctgtccc atatattata ccagttgctt ctgagcccg cagatgcca g cgatgcca	3480
ggaaacaagt tgaaatccca ggaatctctt taactgattt tgctaaaaat c tccctgtga	3540

-continued

```

gccttccact caactcttaa tatgcttgca ttgtttaagt ttttaaattc t gaaaattaa 3600
taattagggg ttttttcata tgtgttgcat aatgcaaacc tcctagggta a aatagtttc 3660
tttatttaag atagaataat ttccagaaat tgtacttttg aggtatcatt t ttatctgta 3720
atggtttgtc tgtctttttt cctctgatca gtattttttt ataccagttt t ggagactgc 3780
ctgagatgaa aggaaatgtg gaataaaagg aggttttctt gatgtgggtg a aagaaaaca 3840
gattccaaga gaattgaaga ttttttttgt ttccttggtg ctttttttct t ttaaattag 3900
gactaatggt tcttttgggg tgcttgaggg atattcatat aaccaaagtt t gagaactgg 3960
gaacttcatg ctgatttgta catattgaag tttctctggt attcaaaggt t ataatagta 4020
atgaattttc attaataaat cactttgtca gaaaaaaaa aaaaaaaaa 4069
    
```

```

<210> SEQ ID NO 4
<211> LENGTH: 976
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
    
```

<400> SEQUENCE: 4

```

Met Asn Ala Gly Pro Gly Cys Glu Pro Cys T hr Lys Ala Thr Arg Trp
1          5          10          15

Gly Ala Ala Thr Thr Ser Pro Ala Ala Ser A sp Ala Arg Ser Phe Pro
          20          25          30

Ser Arg Gln Arg Arg Val Leu Asp Pro Lys A sp Ala His Val Gln Phe
          35          40          45

Arg Val Pro Pro Ser Ser Pro Ala Cys Val P ro Gly Gln Ala Gly Gln
          50          55          60

His Arg Gly Ser Ala Thr Ser Leu Val Phe L ys Gln Lys Thr Ile Thr
65          70          75          80

Ser Trp Met Asp Thr Lys Gly Ile Lys Thr A la Glu Ser Glu Ser Leu
          85          90          95

Asp Ser Lys Glu Asn Asn Asn Thr Arg Ile G lu Ser Met Met Ser Ser
          100          105          110

Val Gln Lys Asp Asn Phe Tyr Gln His Asn V al Glu Lys Leu Val Asn
          115          120          125

Val Ser Gln Leu Ser Leu Asp Lys Ser Leu T hr Glu Lys Ser Thr Gln
          130          135          140

Tyr Leu Asn Gln His Gln Thr Ala Ala Met C ys Lys Trp Gln Asn Glu
145          150          155          160

Gly Lys His Thr Glu Gln Leu Leu Glu Ser G lu Pro Gln Thr Val Thr
          165          170          175

Leu Val Pro Glu Gln Phe Ser Asn Ala Asn I le Asp Arg Ser Pro Gln
          180          185          190

Asn Asp Asp His Ser Asp Thr Asp Ser Glu G lu Asn Arg Asp Asn Gln
          195          200          205

Gln Phe Leu Thr Thr Val Lys Leu Ala Asn A la Lys Gln Thr Thr Glu
          210          215          220

Asp Glu His Ala Arg Glu Ala Lys Ser His G ln Lys Cys Ser Lys Ser
225          230          235          240

Cys His Pro Gly Glu Asp Cys Ala Ser Cys G ln Gln Asp Glu Ile Asp
          245          250          255

Val Val Pro Lys Ser Pro Leu Ser Asp Val G ly Ser Glu Asp Val Gly
          260          265          270

Thr Gly Ser Lys Asn Asp Asn Lys Leu Ile A rg Gln Glu Ser Cys Leu
    
```


-continued

275					280					285					
Gly	Asn	Ser	Pro	Pro	Phe	Glu	Lys	Glu	Ser	Glu	Pro	Glu	Ser	Pro	Met
290						295					300				
Asp	Val	Asp	Asn	Ser	Lys	Asn	Ser	Cys	Gln	Asp	Ser	Glu	Ala	Asp	Glu
305					310					315					320
Glu	Thr	Ser	Pro	Gly	Phe	Asp	Glu	Gln	Glu	Asp	Gly	Ser	Ser	Ser	Gln
				325					330					335	
Thr	Ala	Asn	Lys	Pro	Ser	Arg	Phe	Gln	Ala	Arg	Asp	Ala	Asp	Ile	Glu
			340					345					350		
Phe	Arg	Lys	Arg	Tyr	Ser	Thr	Lys	Gly	Gly	Glu	Val	Arg	Leu	His	Phe
		355					360						365		
Gln	Phe	Glu	Gly	Gly	Glu	Ser	Arg	Thr	Gly	Met	Asn	Asp	Leu	Asn	Ala
	370					375					380				
Lys	Leu	Pro	Gly	Asn	Ile	Ser	Ser	Leu	Asn	Val	Glu	Cys	Arg	Asn	Ser
385					390					395					400
Lys	Gln	His	Gly	Lys	Lys	Asp	Ser	Lys	Ile	Thr	Asp	His	Leu	Met	Arg
				405					410					415	
Leu	Pro	Lys	Ala	Glu	Asp	Arg	Arg	Lys	Glu	Gln	Trp	Glu	Thr	Lys	His
			420					425						430	
Gln	Arg	Thr	Glu	Arg	Lys	Ile	Pro	Lys	Tyr	Val	Pro	Pro	His	Leu	Ser
			435				440						445		
Pro	Asp	Lys	Lys	Trp	Leu	Gly	Thr	Pro	Ile	Glu	Glu	Met	Arg	Arg	Met
	450					455					460				
Pro	Arg	Cys	Gly	Ile	Arg	Leu	Pro	Leu	Leu	Arg	Pro	Ser	Ala	Asn	His
465					470					475					480
Thr	Val	Thr	Ile	Arg	Val	Asp	Leu	Leu	Arg	Ala	Gly	Glu	Val	Pro	Lys
				485					490					495	
Pro	Phe	Pro	Thr	His	Tyr	Lys	Asp	Leu	Trp	Asp	Asn	Lys	His	Val	Lys
			500					505					510		
Met	Pro	Cys	Ser	Glu	Gln	Asn	Leu	Tyr	Pro	Val	Glu	Asp	Glu	Asn	Gly
		515					520					525			
Glu	Arg	Thr	Ala	Gly	Ser	Arg	Trp	Glu	Leu	Ile	Gln	Thr	Ala	Leu	Leu
						535					540				
Asn	Lys	Phe	Thr	Arg	Pro	Gln	Asn	Leu	Lys	Asp	Ala	Ile	Leu	Lys	Tyr
545					550					555					560
Asn	Val	Ala	Tyr	Ser	Lys	Lys	Trp	Asp	Phe	Thr	Ala	Leu	Ile	Asp	Phe
				565					570					575	
Trp	Asp	Lys	Val	Leu	Glu	Glu	Ala	Glu	Ala	Gln	His	Leu	Tyr	Gln	Ser
			580					585						590	
Ile	Leu	Pro	Asp	Met	Val	Lys	Ile	Ala	Leu	Cys	Leu	Pro	Asn	Ile	Cys
	595						600						605		
Thr	Gln	Pro	Ile	Pro	Leu	Leu	Lys	Gln	Lys	Met	Asn	His	Ser	Ile	Thr
	610						615					620			
Met	Ser	Gln	Glu	Gln	Ile	Ala	Ser	Leu	Leu	Ala	Asn	Ala	Phe	Phe	Cys
625					630					635					640
Thr	Phe	Pro	Arg	Arg	Asn	Ala	Lys	Met	Lys	Ser	Glu	Tyr	Ser	Ser	Tyr
				645					650					655	
Pro	Asp	Ile	Asn	Phe	Asn	Arg	Leu	Phe	Glu	Gly	Arg	Ser	Ser	Arg	Lys
			660					665						670	
Pro	Glu	Lys	Leu	Lys	Thr	Leu	Phe	Cys	Tyr	Phe	Arg	Arg	Val	Thr	Glu
		675					680						685		
Lys	Lys	Pro	Thr	Gly	Leu	Val	Thr	Phe	Thr	Arg	Gln	Ser	Leu	Glu	Asp
	690						695				700				

-continued

Phe Pro Glu Trp Glu Arg Cys Glu Lys Pro L eu Thr Arg Leu His Val
 705 710 715 720
 Thr Tyr Glu Gly Thr Ile Glu Glu Asn Gly G ln Gly Met Leu Gln Val
 725 730 735
 Asp Phe Ala Asn Arg Phe Val Gly Gly V al Thr Ser Ala Gly Leu
 740 745 750
 Val Gln Glu Ile Arg Phe Leu Ile Asn P ro Glu Leu Ile Ile Ser
 755 760 765
 Arg Leu Phe Thr Glu Val Leu Asp His Asn G lu Cys Leu Ile Ile Thr
 770 775 780
 Gly Thr Glu Gln Tyr Ser Glu Tyr Thr Gly T yr Ala Glu Thr Tyr Arg
 785 790 795 800
 Trp Ser Arg Ser His Glu Asp Gly Ser Glu A rg Asp Asp Cys Glu Arg
 805 810 815
 Arg Cys Thr Glu Ile Val Ala Ile Asp Ala L eu His Phe Arg Arg Tyr
 820 825 830
 Leu Asp Gln Phe Val Pro Glu Lys Met Arg A rg Glu Leu Asn Lys Ala
 835 840 845
 Tyr Cys Gly Phe Leu Arg Pro Gly Val Ser S er Glu Asn Leu Ser Ala
 850 855 860
 Val Ala Thr Gly Asn Trp Gly Cys Gly Ala P he Gly Gly Asp Ala Arg
 865 870 875 880
 Leu Lys Ala Leu Ile Gln Ile Leu Ala Ala A la Ala Ala Glu Arg Asp
 885 890 895
 Val Val Tyr Phe Thr Phe Gly Asp Ser Glu L eu Met Arg Asp Ile Tyr
 900 905 910
 Ser Met His Ile Phe Leu Thr Glu Arg Lys L eu Thr Val Gly Asp Val
 915 920 925
 Tyr Lys Leu Leu Leu Arg Tyr Tyr Asn Glu G lu Cys Arg Asn Cys Ser
 930 935 940
 Thr Pro Gly Pro Asp Ile Lys Leu Tyr Pro P he Ile Tyr His Ala Val
 945 950 955 960
 Glu Ser Cys Ala Glu Thr Ala Asp His Ser G ly Gln Arg Thr Gly Thr
 965 970 975

<210> SEQ ID NO 5
 <211> LENGTH: 3814
 <212> TYPE: DNA
 <213> ORGANISM: Mus musculus
 <220> FEATURE:
 <400> SEQUENCE: 5

gggggactgt gtgctgcggg tcccagcatg agtgccggcc cgggctggga g ccctgcacg 60
 aaagcgcgct ggggcgccgc tggaacttct ggcgcgactg cctcggactc c cggagcttc 120
 cctggcaggc agaggcgtgt tctcgacccc aaggacgctc ccgccagtt c agggtcctc 180
 ccgtcctcgc cagcctcgt ctcggggcgg gggggaccgc acagaggcaa c gccacctcg 240
 tttgttttca acaaaaagac tattactact tggatggata ctaaaggacc c aagacagct 300
 gaatcagaaa gtaaagaaaa caacaataca agaattgact ccatgatgag t tctgtgcag 360
 aaagataact tttaccaca taaggtgaa aaattgaaa atgttcctca g ctaaattct 420
 gataaatcac ccacagaaaa gagttcacag tatttgaacc aacagcagac t gcgagtgtg 480
 tgcaagtggc agaataaagg gaagcatgca gaacagcttt tggcaagtga g cctcccgcg 540

-continued

gggactccgc	taccaaagca	gcttagtaat	gctaacattg	gtcagtcacc	c cacactgat	600
gaccacagtg	acacagatca	tgaagaagac	agagacaatc	agcagtttct	t acacctata	660
aaacttgcaa	atacaaagcc	aacagtagga	gatgggcagg	ccagaagcaa	c tgtaagtgc	720
agtggatctc	gccagtctgt	gaaagactgt	acaggctgtc	aacaggagga	g gtggatgtg	780
ctaccagaga	gtcctttgtc	agatgttggt	gccgaggaca	ttggaactgg	a ccaaaaaat	840
gacaacaaat	tgactggaca	agaaagcagc	ctaggtgatt	cgctccatt	t gagaaagaa	900
agtgagcctg	agtcaccaat	ggatgtgagc	aactcgagaa	acagttgtca	a gattcagaa	960
gcagatgaag	aaacaagtcc	agtctttgat	gagcaagatg	atcgttctc	c caaacagca	1020
aataaaacttt	caagttgcca	agcaagagaa	gctgatggcg	atcttaggaa	a cggatattg	1080
actaagggaa	gtgaagttag	attgcatttc	caatttgaag	gagaaaataa	t gctgggacc	1140
agtgacttaa	atgccaagcc	atctggaaac	tcttctagcc	ttaatgtaga	g tgtagaagt	1200
tccaagcagc	atggaaaaag	ggattctaaa	attacagatc	atttcatgag	a atttccaag	1260
tcagaggaca	gaagaaaaga	acaatgtgaa	gtcagacatc	aaagaacaga	a aggaagatt	1320
ccaaaaataca	tcccacctaa	cctccctcca	gagaagaagt	ggctgggaac	t cctattgag	1380
gaaatgagaa	aaatgcctcg	gtgtggggtc	catttgcctt	ccttaagacc	a tctgcaagt	1440
cacacagtga	ctgttcgggt	agacctctg	agagcaggag	aggttccgaa	a ccttttcca	1500
acacattaca	aagatttgtg	ggataacaaa	catgtgaaaa	tgcttgttc	g gaacaaaac	1560
ttgtaccctg	tggaaagatg	gaatggtgag	cgaactgcag	ggagttagtg	g gagctcatt	1620
cagactgcac	ttctcaacaa	attcacacga	cccagaact	tgaaggatgc	g attctgaaa	1680
tacaatgtgg	catattctaa	gaaatgggac	tttacagctt	tggttgattt	c tgggataag	1740
gtacttgaag	aagcagaggc	ccaacattta	tatcagtcca	ttttacctga	c atggtgaaa	1800
attgcactct	gtctgccaaa	tatttgcacc	cagccaatac	cactcctgaa	a cagaagatg	1860
aatcattctg	tcacgatgtc	acaggaacag	atcgccagtc	ttttagctaa	t gctttcttc	1920
tgcacatttc	cccgacggaa	tgccaagatg	aaatcggagt	attctagtta	c ccagacatt	1980
aacttcaatc	ggttgtttga	aggacgttca	tcaaggaaac	cagaaaaact	g aaaacactc	2040
ttctgctact	ttcgaagagt	cacagagaaa	aaacctacag	gatttggtag	a tttacaaga	2100
cagagtcttg	aagattttcc	agaatgggaa	agggtgtgaa	agcctctgac	a cgcttacac	2160
gtcacttacg	agggtaccat	agaaggcaac	ggccgaggca	tgctacaggt	g gattttgca	2220
aatcgttttg	ttggaggtg	tgtgactggt	gcgggacttg	tacaagaaga	a atcagattt	2280
ttaatcaatc	ctgaattgat	tgtttcacgg	ctgttctactg	agggtctgga	t cacaatgag	2340
tgtcttatta	tcacaggtac	tgaacagtac	agtgaataca	caggctatgc	t gaaacttat	2400
cgttgggccc	gaagccatga	agatgggagt	gaaaaggacg	attggcagcg	g cgctgcacg	2460
gagatcgttg	ccattgacgc	acttcacttc	agacgtacc	tcgatcagtt	t gtgctgag	2520
aaagtgagac	gtgagcttaa	caaggttac	tgccgattcc	tccgtcctgg	a gttccttct	2580
gaaaatcttt	ctgcagtggc	cacgggaaac	tggggctgtg	gtgccttttg	g ggtgacgct	2640
agattaaaag	ccttaataca	gatcctggca	gctgctgcgg	ctgaacgtga	c gtggtttat	2700
ttcacctttg	gggactcaga	gttgatgaga	gacatttaca	gcatgcacac	t ttccttacc	2760
gagagggaagc	tggatgttg	aaaagtgtac	aagttattgc	ttagatacta	c aatgaagaa	2820
tgacagaaact	gttccacccc	tggaccagac	atcaagcttt	atccattcat	a taccatgct	2880
gttagtcaa	gtgcagagac	cactgacatg	ccaggacaga	aggcaggcac	c tgaggaaca	2940

-continued

```

agtgactagg acctcctctc aaagagacat cctatttgaa atgtggggtg t gatgtctga 3000
attgactgaa tctgatctaa gtgtgtatat aatccacatt tgtaatcaag g atgcagtct 3060
cttctgcata tgcagttggt tcttggtcat cctgggtggac atgcctttag a catggcttc 3120
ttcaattttt cttctccttc agtctttatt ctttgatttt ttttttccaa c ttgatttct 3180
tgggaaaaat caagaaagt tgcactcagc ttctagatct ttctcttct g tctgtgtgt 3240
tgtccagact gctttggtgg cttagcagata ccatcacact tggaggaagt t acaaatcca 3300
gaaatctgag tttgctgcag atttacctgt gagcttctca ctccaaccc t tgttaggct 3360
tgtgtgtct acattttcaa ttttggaaagt tgaagttttt cttatgttac t taatgctag 3420
tatcttttag gctaaaacta ttttctattt aaggcagact aatttccagt t tctcttttg 3480
aaacatcatc cctataagta acgggttttt tctgctttt ttcccagcg c tattttaga 3540
agctggccaa gaggaaagaa aatgtagaat aaaaggattt tcctcggatg c tataaagaa 3600
gccaggttca agagcgttgg ggtttttgg tttttcaaga cttgtttttc c tttgcagct 3660
agggtgagtg cttgttctgt ggtgctgagg gcatagtct gtaaccaaag g tctttgctg 3720
gagacttgat gctgatttgt acatatggaa gtttctctgg caggaaatat t agagttaat 3780
aaatttcatt aataaatcat ttgtcagaaa aaaa 3814

```

```

<210> SEQ ID NO 6
<211> LENGTH: 968
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:

```

<400> SEQUENCE: 6

```

Met Ser Ala Gly Pro Gly Trp Glu Pro Cys T hr Lys Ala Arg Trp Gly
1 5 10 15
Ala Ala Gly Thr Ser Ala Pro Thr Ala Ser A sp Ser Arg Ser Phe Pro
20 25 30
Gly Arg Gln Arg Arg Val Leu Asp Pro Lys A sp Ala Pro Val Gln Phe
35 40 45
Arg Val Pro Pro Ser Ser Pro Ala Cys Val S er Gly Arg Ala Gly Pro
50 55 60
His Arg Gly Asn Ala Thr Ser Phe Val Phe L ys Gln Lys Thr Ile Thr
65 70 75 80
Thr Trp Met Asp Thr Lys Gly Pro Lys Thr A la Glu Ser Glu Ser Lys
85 90 95
Glu Asn Asn Asn Thr Arg Ile Asp Ser Met M et Ser Ser Val Gln Lys
100 105 110
Asp Asn Phe Tyr Pro His Lys Val Glu Lys L eu Glu Asn Val Pro Gln
115 120 125
Leu Asn Leu Asp Lys Ser Pro Thr Glu Lys S er Ser Gln Tyr Leu Asn
130 135 140
Gln Gln Gln Thr Ala Ser Val Cys Lys Trp G ln Asn Glu Gly Lys His
145 150 155 160
Ala Glu Gln Leu Leu Ala Ser Glu Pro Pro A la Gly Thr Pro Leu Pro
165 170 175
Lys Gln Leu Ser Asn Ala Asn Ile Gly Gln S er Pro His Thr Asp Asp
180 185 190
His Ser Asp Thr Asp His Glu Glu Asp Arg A sp Asn Gln Gln Phe Leu
195 200 205

```

-continued

Thr Pro Ile Lys Leu Ala Asn Thr Lys Pro T hr Val Gly Asp Gly Gln
 210 215 220

Ala Arg Ser Asn Cys Lys Cys Ser Gly Ser A rg Gln Ser Val Lys Asp
 225 230 235 240

Cys Thr Gly Cys Gln Gln Glu Glu Val Asp V al Leu Pro Glu Ser Pro
 245 250 255

Leu Ser Asp Val Gly Ala Glu Asp Ile Gly T hr Gly Pro Lys Asn Asp
 260 265 270

Asn Lys Leu Thr Gly Gln Glu Ser Ser Leu G ly Asp Ser Pro Pro Phe
 275 280 285

Glu Lys Glu Ser Glu Pro Glu Ser Pro Met A sp Val Asp Asn Ser Arg
 290 295 300

Asn Ser Cys Gln Asp Ser Glu Ala Asp Glu G lu Thr Ser Pro Val Phe
 305 310 315 320

Asp Glu Gln Asp Asp Arg Ser Ser Gln Thr A la Asn Lys Leu Ser Ser
 325 330 335

Cys Gln Ala Arg Glu Ala Asp Gly Asp Leu A rg Lys Arg Tyr Leu Thr
 340 345 350

Lys Gly Ser Glu Val Arg Leu His Phe Gln P he Glu Gly Glu Asn Asn
 355 360 365

Ala Gly Thr Ser Asp Leu Asn Ala Lys Pro S er Gly Asn Ser Ser Ser
 370 375 380

Leu Asn Val Glu Cys Arg Ser Ser Lys Gln H is Gly Lys Arg Asp Ser
 385 390 395 400

Lys Ile Thr Asp His Phe Met Arg Ile Ser L ys Ser Glu Asp Arg Arg
 405 410 415

Lys Glu Gln Cys Glu Val Arg His Gln Arg T hr Glu Arg Lys Ile Pro
 420 425 430

Lys Tyr Ile Pro Pro Asn Leu Pro Pro Glu L ys Lys Trp Leu Gly Thr
 435 440 445

Pro Ile Glu Glu Met Arg Lys Met Pro Arg C ys Gly Ile His Leu Pro
 450 455 460

Ser Leu Arg Pro Ser Ala Ser His Thr Val T hr Val Arg Val Asp Leu
 465 470 475 480

Leu Arg Ala Gly Glu Val Pro Lys Pro Phe P ro Thr His Tyr Lys Asp
 485 490 495

Leu Trp Asp Asn Lys His Val Lys Met Pro C ys Ser Glu Gln Asn Leu
 500 505 510

Tyr Pro Val Glu Asp Glu Asn Gly Glu Arg T hr Ala Gly Ser Arg Trp
 515 520 525

Glu Leu Ile Gln Thr Ala Leu Leu Asn Lys P he Thr Arg Pro Gln Asn
 530 535 540

Leu Lys Asp Ala Ile Leu Lys Tyr Asn Val A la Tyr Ser Lys Lys Trp
 545 550 555 560

Asp Phe Thr Ala Leu Val Asp Phe Trp Asp L ys Val Leu Glu Glu Ala
 565 570 575

Glu Ala Gln His Leu Tyr Gln Ser Ile Leu P ro Asp Met Val Lys Ile
 580 585 590

Ala Leu Cys Leu Pro Asn Ile Cys Thr Gln P ro Ile Pro Leu Leu Lys
 595 600 605

Gln Lys Met Asn His Ser Val Thr Met Ser G ln Glu Gln Ile Ala Ser
 610 615 620

Leu Leu Ala Asn Ala Phe Phe Cys Thr Phe P ro Arg Arg Asn Ala Lys

-continued

625	630	635	640
Met Lys Ser Glu Tyr Ser Ser Tyr Pro Asp I le Asn Phe Asn Arg Leu	645	650	655
Phe Glu Gly Arg Ser Ser Arg Lys Pro Glu L ys Leu Lys Thr Leu Phe	660	665	670
Cys Tyr Phe Arg Arg Val Thr Glu Lys Lys P ro Thr Gly Leu Val Thr	675	680	685
Phe Thr Arg Gln Ser Leu Glu Asp Phe Pro G lu Trp Glu Arg Cys Glu	690	695	700
Lys Pro Leu Thr Arg Leu His Val Thr Tyr G lu Gly Thr Ile Glu Gly	705	710	720
Asn Gly Arg Gly Met Leu Gln Val Asp Phe A la Asn Arg Phe Val Gly	725	730	735
Gly Gly Val Thr Gly Ala Gly Leu Val Gln G lu Glu Ile Arg Phe Leu	740	745	750
Ile Asn Pro Glu Leu Ile Val Ser Arg Leu P he Thr Glu Val Leu Asp	755	760	765
His Asn Glu Cys Leu Ile Ile Thr Gly Thr G lu Gln Tyr Ser Glu Tyr	770	775	780
Thr Gly Tyr Ala Glu Thr Tyr Arg Trp Ala A rg Ser His Glu Asp Gly	785	790	800
Ser Glu Lys Asp Asp Trp Gln Arg Arg Cys T hr Glu Ile Val Ala Ile	805	810	815
Asp Ala Leu His Phe Arg Arg Tyr Leu Asp G ln Phe Val Pro Glu Lys	820	825	830
Val Arg Arg Glu Leu Asn Lys Ala Tyr Cys G ly Phe Leu Arg Pro Gly	835	840	845
Val Pro Ser Glu Asn Leu Ser Ala Val Ala T hr Gly Asn Trp Gly Cys	850	855	860
Gly Ala Phe Gly Gly Asp Ala Arg Leu Lys A la Leu Ile Gln Ile Leu	865	870	880
Ala Ala Ala Ala Ala Glu Arg Asp Val Val T yr Phe Thr Phe Gly Asp	885	890	895
Ser Glu Leu Met Arg Asp Ile Tyr Ser Met H is Thr Phe Leu Thr Glu	900	905	910
Arg Lys Leu Asp Val Gly Lys Val Tyr Lys L eu Leu Leu Arg Tyr Tyr	915	920	925
Asn Glu Glu Cys Arg Asn Cys Ser Thr Pro G ly Pro Asp Ile Lys Leu	930	935	940
Tyr Pro Phe Ile Tyr His Ala Val Glu Ser S er Ala Glu Thr Thr Asp	945	950	955
Met Pro Gly Gln Lys Ala Gly Thr	965		

<210> SEQ ID NO 7
 <211> LENGTH: 2781
 <212> TYPE: DNA
 <213> ORGANISM: Drosophila melanogaster
 <220> FEATURE:

<400> SEQUENCE: 7

tcgaaagtgtg tggtatttat aaagtgcgat attcatcaca gctatcgctc a tccccaaaa	60
caccggtagt caagaattca ggtcacactt gatttttccg atattccaaa a ggtttacca	120
atctacggca aatcgccgca gagcaagtgc atccgtgctg accaatcgac t cggaaggc	180

-continued

tttgtgctta aactgcgcca ggatgtcgaa gtcgccgat ggcgggattt c cgaataga	240
aacggaggag gagccgaaa atctggcgaa ctccctagat gattcgtggc g tggagtttc	300
catggaggct atacatcgta atcggcagcc tttcgaattg gagaatttgc c accagtgc	360
tgccggcaat ctccaccggg ttatgtacca gctgccaatt cgtgaaacac c gccaccgcc	420
ctacaaatca cgggaaagt gggactccga gcatgtgcgt ctgccctgtg c gcccgagtc	480
gaaatatccg agggagaatc cggatggcag caccaccatc gattttcgct g ggaatgat	540
cgaacgagcc ctctgcagc ccataaagac gtgtgaggaa ctgcaggcgg c gataatgc	600
atataatacc acctataggg atcagtggca ctttcgtgcc ctccatcaac t tctcgacga	660
ggaactggac gagagcgaaa cacgggtttt ctccgaggat ctattgccgc g cattatccg	720
attggcattg cggctaccgg acttgattca atcgcagtt ccgctgctca a gccaccaca	780
gaacgcctca ttgagctga gccaacagca gatctcctgc ctgttgcca a tgccttctt	840
gtgcacgttt ccccgaaaga acaccctcaa gaggaagtcc gagtacagca c ttttccaga	900
catcaacttt aacaggcttt accaatcgac gggaccggca gttctggaga a gcttaaatg	960
cattatgcac tattttcgtc gcgtgtgtcc cacagagcgg gatgccagca a tgtgccac	1020
cgggtgtgta acctttgttc gtcggagcgg attgccgaa catctgatcg a ctggagcca	1080
aagtgcggcg ccgttgggtg atgtgccatt gcacgtgat gccgagggaa c aatcgagga	1140
tgagggcatt ggactgtcgc aagtagactt tgccaacaaa tatttgggtg g cgggtgtctt	1200
gggacatgac tgcgttcagc aggagatacg ctttgttacc tgtccggagc t attggtggg	1260
taaaactctt acggagtgtc tgcgaccatt cgaggccctg gtgatgttg g cgccgaaag	1320
gtatagtaac tatacgggat atgccggaag ctccgagtgg tccggcaact t tgaggattc	1380
aacccaaga gatagctcag gtcgtcgaca aacggccatt gtggcaatcg a tgcctaca	1440
ttttgccag tcacatcacc aatctcgcga ggatctcatg gaaagggagc t gaacaaggc	1500
gtacattgga tttgttccact ggatggtgac gccgccaccg ggtgtggcaa c tggtaactg	1560
gggttgcggc gcattcggcg gtgactccta tctgaaagcc ctgctgcaac t tatggtctg	1620
cgccagttg ggcagacctt tggcctacta tacctttgga aatgtggagt t tagggatga	1680
ttttcatgaa atgtggctgt tgtttcgaaa tgacgggact acggtgcagc a gctttggag	1740
tattttaagg tcgtacagta ggcttattaa ggagaagagc tccaaggagc c gcgtgagaa	1800
taaggcatcc aaaaagaagc tatatgattt tattaagag gaacttaaga a ggtcagaga	1860
gtgtcccga gagggagcat ccgccgaagc tggaaagtct agagttagctg g attagcga	1920
agggaaatca gaaacatcag cgaatcctc gccagaactc aacaagcaac c cgccgacc	1980
gcaaatcacc ataacgcaac aaagtaccga tctattgcc cgcgaattat c gcaagataa	2040
ctctaattct tcggaagatc aggcccttct tatgctgtcg gacgatgag a ggccaatgc	2100
catgatggag gccgctatgc tggaggctaa aagcagcgt gaaataagca a cagcagcac	2160
aacgtccaaa acgagcagta cagccacgaa atcaatgggt tcaggtggcc g ccagttgag	2220
tctgctcgag atgtgagca ccattatga aaagggttcg gcctcgaaga g gccacgaaa	2280
atcacccaac tgcagcaag ctgagggttc agcaagagt cgtaaggaga t cgatgtgac	2340
cgacaagac gaaagagac atattgttga ctagggtgata ttgcaactaca g gattgttac	2400
tgccccaaa aattgaagag gtataaatg tattgtagat aactttaag a catatttag	2460
ggcattttaa agtaggatca ttgtaagtgc aataaagtga aatTTTTTT t tttttaa	2520

-continued

tatactattc taatctgcaa agacaatttt actgttaa at ttgtataaca t tcgaattaa 2580
 ttaataataat ttgttatatc atgcaaatct agcttttatt atgcgaaatt t gtagttaa 2640
 gccagtaaag tttcttttta ttaaccgaa accttttggt tatttttatt g accacaaca 2700
 agaacatcaa caacaacaac cacgaaaaa aagcgaatat atatttggtt g ttcgtatat 2760
 atatatatat ctaagcagat c 2781

<210> SEQ ID NO 8
 <211> LENGTH: 768
 <212> TYPE: PRT
 <213> ORGANISM: Drosophila melanogaster
 <220> FEATURE:

<400> SEQUENCE: 8

Met Gln Glu Phe Arg Ser His Leu Ile Phe P ro Ile Phe Gln Lys Val
 1 5 10 15
 Tyr Gln Ser Thr Ala Asn Arg Arg Arg Ala S er Ala Ser Val Leu Thr
 20 25 30
 Asn Arg Leu Gly Lys Ala Leu Cys Leu Asn C ys Ala Arg Met Ser Lys
 35 40 45
 Ser Pro Asp Gly Gly Ile Ser Glu Ile Glu T hr Glu Glu Glu Pro Glu
 50 55 60
 Asn Leu Ala Asn Ser Leu Asp Asp Ser Trp A rg Gly Val Ser Met Glu
 65 70 75 80
 Ala Ile His Arg Asn Arg Gln Pro Phe Glu L eu Glu Asn Leu Pro Pro
 85 90 95
 Val Thr Ala Gly Asn Leu His Arg Val Met T yr Gln Leu Pro Ile Arg
 100 105 110
 Glu Thr Pro Pro Arg Pro Tyr Lys Ser Pro G ly Lys Trp Asp Ser Glu
 115 120 125
 His Val Arg Leu Pro Cys Ala Pro Glu Ser L ys Tyr Pro Arg Glu Asn
 130 135 140
 Pro Asp Gly Ser Thr Thr Ile Asp Phe Arg T rp Glu Met Ile Glu Arg
 145 150 155 160
 Ala Leu Leu Gln Pro Ile Lys Thr Cys Glu G lu Leu Gln Ala Ala Ile
 165 170 175
 Ile Ser Tyr Asn Thr Thr Tyr Arg Asp Gln T rp His Phe Arg Ala Leu
 180 185 190
 His Gln Leu Leu Asp Glu Glu Leu Asp Glu S er Glu Thr Arg Val Phe
 195 200 205
 Phe Glu Asp Leu Leu Pro Arg Ile Ile Arg L eu Ala Leu Arg Leu Pro
 210 215 220
 Asp Leu Ile Gln Ser Pro Val Pro Leu Leu L ys His His Lys Asn Ala
 225 230 235 240
 Ser Leu Ser Leu Ser Gln Gln Gln Ile Ser C ys Leu Leu Ala Asn Ala
 245 250 255
 Phe Leu Cys Thr Phe Pro Arg Arg Asn Thr L eu Lys Arg Lys Ser Glu
 260 265 270
 Tyr Ser Thr Phe Pro Asp Ile Asn Phe Asn A rg Leu Tyr Gln Ser Thr
 275 280 285
 Gly Pro Ala Val Leu Glu Lys Leu Lys Cys I le Met His Tyr Phe Arg
 290 295 300
 Arg Val Cys Pro Thr Glu Arg Asp Ala Ser A sn Val Pro Thr Gly Val
 305 310 315 320

-continued

Val Thr Phe Val Arg Arg Ser Gly Leu Pro G lu His Leu Ile Asp Trp
 325 330 335

Ser Gln Ser Ala Ala Pro Leu Gly Asp Val P ro Leu His Val Asp Ala
 340 345 350

Glu Gly Thr Ile Glu Asp Glu Gly Ile Gly L eu Leu Gln Val Asp Phe
 355 360 365

Ala Asn Lys Tyr Leu Gly Gly Gly Val Leu G ly His Gly Cys Val Gln
 370 375 380

Glu Glu Ile Arg Phe Val Ile Cys Pro Glu L eu Leu Val Gly Lys Leu
 385 390 395 400

Phe Thr Glu Cys Leu Arg Pro Phe Glu Ala L eu Val Met Leu Gly Ala
 405 410 415

Glu Arg Tyr Ser Asn Tyr Thr Gly Tyr Ala G ly Ser Phe Glu Trp Ser
 420 425 430

Gly Asn Phe Glu Asp Ser Thr Pro Arg Asp S er Ser Gly Arg Arg Gln
 435 440 445

Thr Ala Ile Val Ala Ile Asp Ala Leu His P he Ala Gln Ser His His
 450 455 460

Gln Tyr Arg Glu Asp Leu Met Glu Arg Glu L eu Asn Lys Ala Tyr Ile
 465 470 475 480

Gly Phe Val His Trp Met Val Thr Pro Pro P ro Gly Val Ala Thr Gly
 485 490 495

Asn Trp Gly Cys Gly Ala Phe Gly Gly Asp S er Tyr Leu Lys Ala Leu
 500 505 510

Leu Gln Leu Met Val Cys Ala Gln Leu Gly A rg Pro Leu Ala Tyr Tyr
 515 520 525

Thr Phe Gly Asn Val Glu Phe Arg Asp Asp P he His Glu Met Trp Leu
 530 535 540

Leu Phe Arg Asn Asp Gly Thr Thr Val Gln G ln Leu Trp Ser Ile Leu
 545 550 555 560

Arg Ser Tyr Ser Arg Leu Ile Lys Glu Lys S er Ser Lys Glu Pro Arg
 565 570 575

Glu Asn Lys Ala Ser Lys Lys Lys Leu Tyr A sp Phe Ile Lys Glu Glu
 580 585 590

Leu Lys Lys Val Arg Asp Val Pro Gly Glu G ly Ala Ser Ala Glu Ala
 595 600 605

Gly Ser Ser Arg Val Ala Gly Leu Gly Glu G ly Lys Ser Glu Thr Ser
 610 615 620

Ala Lys Ser Ser Pro Glu Leu Asn Lys Gln P ro Ala Arg Pro Gln Ile
 625 630 635 640

Thr Ile Thr Gln Gln Ser Thr Asp Leu Leu P ro Ala Gln Leu Ser Gln
 645 650 655

Asp Asn Ser Asn Ser Ser Glu Asp Gln Ala L eu Leu Met Leu Ser Asp
 660 665 670

Asp Glu Glu Ala Asn Ala Met Met Glu Ala A la Ser Leu Glu Ala Lys
 675 680 685

Ser Ser Val Glu Ile Ser Asn Ser Ser Thr T hr Ser Lys Thr Ser Ser
 690 695 700

Thr Ala Thr Lys Ser Met Gly Ser Gly Gly A rg Gln Leu Ser Leu Leu
 705 710 715 720

Glu Met Leu Asp Thr His Tyr Glu Lys Gly S er Ala Ser Lys Arg Pro
 725 730 735

Arg Lys Ser Pro Asn Cys Ser Lys Ala Glu G ly Ser Ala Lys Ser Arg

-continued

740	745	750	
Lys Glu Ile Asp Val Thr Asp Lys Asp Glu L ys Asp Asp Ile Val Asp			
755	760	765	
<210> SEQ ID NO 9 <211> LENGTH: 2181 <212> TYPE: DNA <213> ORGANISM: Caenorhabditis elegans <220> FEATURE:			
<400> SEQUENCE: 9			
atgagcaaga agtttatcga actgggtgat cctgtcactc aagacgagaa a gactacgaa			60
gactatgtcg gagtgtgttt cgcgcatcaa gtcccgacaa tgaaaaggcg g aagttgaca			120
gaacatggaa atactacaga atcaaaagaa gatcctgaag agccaaaaag c cgtgacgta			180
tttgtctcct cgcagtcaag tgatgagagt caagaagatt cggctgaaaa t cgggagatc			240
gctaagaag tgtcagaaaa ttgtgaaat ctgacagaaa ctctcaaaat t tctaataat			300
gagagtttgg acaatgttac tgaagatct gaacacactc ttgataatca c aaaagtact			360
gaaccaatgg aagaagatgt aaacaacaag tccaatattg acgttgcgat t aattctgac			420
gaggatgat aacttgttct ggaagagaat aataaagaaa tgagggatgg a gaacaagta			480
caacagttgt cacaggattt attcgtgat gatcaagagc taattgaata t ccaggaatt			540
atgaaagaca ctacaactca actggatata acagattctg aagtggagac t gctcaaaaa			600
atggaaatga ttgaagaac tgaagcagat tcgacatttg taggcgagga t tcaaaagct			660
acgaaaactg tgaggacatc cagttcaagt ttcctgtcaa ctgtttcaac a tgcgaagcc			720
cctgcaaaag gacgagcaag aatgtatcaa aaagagttgg aaaagcatgt g attgcattt			780
actgagggaa atctcacact acaaccagat ttgaacaaag ttgatcccga c agaaactat			840
cgatattgta caattccgaa ctttccagct tccaaggaa aacttcgaga a gataatcga			900
tatggcccaa aaatcgtttt gcctcaaaga tggcgagaat ttgattcgag g ggccgtaga			960
agagactcat atttctattht caaacgtaag ctcgatggat atttgaaatg c tacaaaaa			1020
actggatatt ttatgtttgt tggacttttg cacacatgt gggaaattga c ccagacatc			1080
acataataac tgccagcact ggaaatgtat tacaagaga tgcggaact t gttggtaga			1140
gaagaggttt tggaaaaatt tgcacgagtt gcccgcatcg caaaaactgc t gaagatatt			1200
ctgccagagc gaatttatcg tcttgttgg gacgtcgaat cagctacctt g agccacaag			1260
caatgtgctg cacttgttgc gagaatgttt tttgcccgac cggacagtcc t ttcagtttc			1320
tgccgaattc tctcgtctga taaatctatt tgtgtggaga aacttaaatt c ctgttcaact			1380
tatttcgaca aaatgtcaat ggatccaccg gatggtgccg tcagtttttag a cttacaaa			1440
atggataaag atacgttcaa cgaagagtg aaagataaaa aattacgttc t ctctctgaa			1500
gttgaattct ttgatgaaat gcttattgaa gacacagctc tctgtacaca a gttgatttt			1560
gcgaacgaac atcttggtag cggagtttta aatcatgggt ctgttcagga g gagatccgt			1620
ttctttagat gtccagaaat gatggttga atgttgtgt gcgagaaaat g aaacaactg			1680
gaagcgattt caattgttgg agcttacgtt ttcagttctt atactggta t ggtcactact			1740
ctaaaatggg cagaacttca accaaatcat tctcgtcaga atacaaacga a tttogagat			1800
cgttttggac gtcttcgggt agaaactatt gcaatcgatg caattctgtt c aaaggatca			1860
aaattagatt gtcagacgga gcagttaaac aaagcaaata tcattagga a atgaagaaa			1920
gcatctatcg gattcatgag ccaggggaccg aaattcacia atattccaat t gttactgga			1980

-continued

tggtggggat gtggagcatt taatggggac aagccactga agttcataat c caagtaatt 2040
 gctgccggag tcgctgatcg tccacttcat ttctgttcat ttggagaacc c gagcttgcc 2100
 gcaaagtgca agaaaattat agaacgaatg aaacagaagg acgtaacact t ggtaagtca 2160
 tgtttttcaa tcttcagttg a 2181

<210> SEQ ID NO 10
 <211> LENGTH: 726
 <212> TYPE: PRT
 <213> ORGANISM: Caenorhabditis elegans
 <220> FEATURE:

<400> SEQUENCE: 10

Met Ser Lys Lys Phe Ile Glu Leu Gly Asp P ro Val Thr Gln Asp Glu
 1 5 10 15
 Lys Asp Tyr Glu Asp Tyr Val Gly Val Gly P he Ala His Gln Val Pro
 20 25 30
 Thr Met Lys Arg Arg Lys Leu Thr Glu His G ly Asn Thr Thr Glu Ser
 35 40 45
 Lys Glu Asp Pro Glu Glu Pro Lys Ser Arg A sp Val Phe Val Ser Ser
 50 55 60
 Gln Ser Ser Asp Glu Ser Gln Glu Asp Ser A la Glu Asn Pro Glu Ile
 65 70 75 80
 Ala Lys Glu Val Ser Glu Asn Cys Glu Asn L eu Thr Glu Thr Leu Lys
 85 90 95
 Ile Ser Asn Ile Glu Ser Leu Asp Asn Val T hr Glu Arg Ser Glu His
 100 105 110
 Thr Leu Asp Asn His Lys Ser Thr Glu Pro M et Glu Glu Asp Val Asn
 115 120 125
 Asn Lys Ser Asn Ile Asp Val Ala Ile Asn S er Asp Glu Asp Asp Glu
 130 135 140
 Leu Val Leu Glu Glu Asn Asn Lys Glu Met A rg Asp Gly Glu Gln Val
 145 150 155 160
 Gln Gln Leu Ser Gln Asp Leu Phe Ala Asp A sp Gln Glu Leu Ile Glu
 165 170 175
 Tyr Pro Gly Ile Met Lys Asp Thr Thr Thr G ln Leu Asp Ile Thr Asp
 180 185 190
 Ser Glu Val Glu Thr Ala Gln Lys Met Glu M et Ile Glu Glu Thr Glu
 195 200 205
 Ala Asp Ser Thr Phe Val Gly Glu Asp Ser L ys Ala Thr Lys Thr Val
 210 215 220
 Arg Thr Ser Ser Ser Ser Phe Leu Ser Thr V al Ser Thr Cys Glu Ala
 225 230 235 240
 Pro Ala Lys Gly Arg Ala Arg Met Tyr Gln L ys Glu Leu Glu Lys His
 245 250 255
 Val Ile Ala Phe Thr Glu Gly Asn Leu Thr L eu Gln Pro Asp Leu Asn
 260 265 270
 Lys Val Asp Pro Asp Arg Asn Tyr Arg Tyr C ys Thr Ile Pro Asn Phe
 275 280 285
 Pro Ala Ser Gln Gly Lys Leu Arg Glu Asp A sn Arg Tyr Gly Pro Lys
 290 295 300
 Ile Val Leu Pro Gln Arg Trp Arg Glu Phe A sp Ser Arg Gly Arg Arg
 305 310 315 320
 Arg Asp Ser Tyr Phe Tyr Phe Lys Arg Lys L eu Asp Gly Tyr Leu Lys

-continued

325										330					335				
Cys	Tyr	Lys	Thr	Thr	Gly	Tyr	Phe	Met	Phe	Val	Gly	Leu	Leu	His	Asn				
			340					345					350						
Met	Trp	Glu	Phe	Asp	Pro	Asp	Ile	Thr	Tyr	Lys	Leu	Pro	Ala	Leu	Glu				
		355					360					365							
Met	Tyr	Tyr	Lys	Glu	Met	Ser	Glu	Leu	Val	Gly	Arg	Glu	Glu	Val	Leu				
	370					375					380								
Glu	Lys	Phe	Ala	Arg	Val	Ala	Arg	Ile	Ala	Lys	Thr	Ala	Glu	Asp	Ile				
385					390					395					400				
Leu	Pro	Glu	Arg	Ile	Tyr	Arg	Leu	Val	Gly	Asp	Val	Glu	Ser	Ala	Thr				
				405					410					415					
Leu	Ser	His	Lys	Gln	Cys	Ala	Ala	Leu	Val	Ala	Arg	Met	Phe	Phe	Ala				
			420					425					430						
Arg	Pro	Asp	Ser	Pro	Phe	Ser	Phe	Cys	Arg	Ile	Leu	Ser	Ser	Asp	Lys				
		435					440					445							
Ser	Ile	Cys	Val	Glu	Lys	Leu	Lys	Phe	Leu	Phe	Thr	Tyr	Phe	Asp	Lys				
	450					455					460								
Met	Ser	Met	Asp	Pro	Pro	Asp	Gly	Ala	Val	Ser	Phe	Arg	Leu	Thr	Lys				
465					470					475					480				
Met	Asp	Lys	Asp	Thr	Phe	Asn	Glu	Glu	Trp	Lys	Asp	Lys	Lys	Leu	Arg				
				485					490					495					
Ser	Leu	Pro	Glu	Val	Glu	Phe	Phe	Asp	Glu	Met	Leu	Ile	Glu	Asp	Thr				
			500					505					510						
Ala	Leu	Cys	Thr	Gln	Val	Asp	Phe	Ala	Asn	Glu	His	Leu	Gly	Gly	Gly				
		515					520					525							
Val	Leu	Asn	His	Gly	Ser	Val	Gln	Glu	Glu	Ile	Arg	Phe	Leu	Met	Cys				
	530					535					540								
Pro	Glu	Met	Met	Val	Gly	Met	Leu	Leu	Cys	Glu	Lys	Met	Lys	Gln	Leu				
545					550					555				560					
Glu	Ala	Ile	Ser	Ile	Val	Gly	Ala	Tyr	Val	Phe	Ser	Ser	Tyr	Thr	Gly				
				565					570					575					
Tyr	Gly	His	Thr	Leu	Lys	Trp	Ala	Glu	Leu	Gln	Pro	Asn	His	Ser	Arg				
			580					585					590						
Gln	Asn	Thr	Asn	Glu	Phe	Arg	Asp	Arg	Phe	Gly	Arg	Leu	Arg	Val	Glu				
		595					600						605						
Thr	Ile	Ala	Ile	Asp	Ala	Ile	Leu	Phe	Lys	Gly	Ser	Lys	Leu	Asp	Cys				
	610					615						620							
Gln	Thr	Glu	Gln	Leu	Asn	Lys	Ala	Asn	Ile	Ile	Arg	Glu	Met	Lys	Lys				
625					630					635					640				
Ala	Ser	Ile	Gly	Phe	Met	Ser	Gln	Gly	Pro	Lys	Phe	Thr	Asn	Ile	Pro				
				645					650					655					
Ile	Val	Thr	Gly	Trp	Trp	Gly	Cys	Gly	Ala	Phe	Asn	Gly	Asp	Lys	Pro				
			660					665					670						
Leu	Lys	Phe	Ile	Ile	Gln	Val	Ile	Ala	Ala	Gly	Val	Ala	Asp	Arg	Pro				
		675					680						685						
Leu	His	Phe	Cys	Ser	Phe	Gly	Glu	Pro	Glu	Leu	Ala	Ala	Lys	Cys	Lys				
	690					695						700							
Lys	Ile	Ile	Glu	Arg	Met	Lys	Gln	Lys	Asp	Val	Thr	Leu	Gly	Lys	Ser				
705					710					715					720				
Cys	Phe	Ser	Ile	Phe	Ser														
				725															

-continued

<211> LENGTH: 31
 <212> TYPE: PRT
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 11

Leu Phe Thr Glu Val Leu Asp His Asn Glu C ys Leu Ile Ile Thr Gly
 1 5 10 15
 Thr Glu Gln Tyr Ser Glu Tyr Thr Gly Tyr A la Glu Thr Tyr Arg
 20 25 30

<210> SEQ ID NO 12
 <211> LENGTH: 29
 <212> TYPE: PRT
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 12

Ala Tyr Cys Gly Phe Leu Arg Pro Gly Val S er Ser Glu Asn Leu Ser
 1 5 10 15
 Ala Val Ala Thr Gly Asn Xaa Gly Cys Gly A la Phe Gly
 20 25

<210> SEQ ID NO 13
 <211> LENGTH: 11
 <212> TYPE: PRT
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 13

Phe Leu Ile Asn Pro Glu Leu Ile Val Ser A rg
 1 5 10

<210> SEQ ID NO 14
 <211> LENGTH: 16
 <212> TYPE: PRT
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 14

Ile Ala Leu Xaa Leu Pro Asn Ile Xaa Thr G ln Pro Ile Pro Leu Leu
 1 5 10 15

<210> SEQ ID NO 15
 <211> LENGTH: 17
 <212> TYPE: DNA
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 15

gaycayaayg artggyt 17

<210> SEQ ID NO 16
 <211> LENGTH: 17
 <212> TYPE: DNA
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 16

ckrtangtyt cngcrta 17

<210> SEQ ID NO 17
 <211> LENGTH: 24
 <212> TYPE: DNA
 <213> ORGANISM: Bos taurus

-continued

<220> FEATURE:

<400> SEQUENCE: 17

atcatcacag gtactgagca gtac 24

<210> SEQ ID NO 18

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 18

gcctgtgtat tcactgtact gctc 24

<210> SEQ ID NO 19

<211> LENGTH: 26

<212> TYPE: PRT

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 19

Glu Asp Lys Arg Lys Glu Gln Cys Glu Met L ys His Gln Arg Thr Glu
1 5 10 15Arg Lys Ile Pro Lys Tyr Ile Pro Pro His
20 25

<210> SEQ ID NO 20

<211> LENGTH: 26

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<220> FEATURE:

<400> SEQUENCE: 20

Glu Asp Arg Arg Lys Glu Gln Trp Glu Thr L ys His Gln Arg Thr Glu
1 5 10 15Arg Lys Ile Pro Lys Tyr Val Pro Pro His
20 25

<210> SEQ ID NO 21

<211> LENGTH: 26

<212> TYPE: PRT

<213> ORGANISM: Mus Musculus

<220> FEATURE:

<400> SEQUENCE: 21

Glu Asp Arg Arg Lys Glu Gln Cys Glu Val A rg His Gln Arg Thr Glu
1 5 10 15Arg Lys Ile Pro Lys Tyr Ile Pro Pro Asn
20 25

<210> SEQ ID NO 22

<211> LENGTH: 32

<212> TYPE: PRT

<213> ORGANISM: Caenorhabditis elegans

<220> FEATURE:

<400> SEQUENCE: 22

His Gln Val Pro Thr Met Lys Arg Arg Lys L eu Thr Glu His Gly Asn
1 5 10 15Thr Thr Glu Ser Leu Leu Leu Lys Glu Asp P ro Pro Glu Pro Lys Ser
20 25 30

<210> SEQ ID NO 23

-continued

<211> LENGTH: 26
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:

<400> SEQUENCE: 23

Glu Gly Lys Arg Lys Gly Asp Glu Val Asp G ly Val Asp Glu Val Ala
 1 5 10 15

Lys Lys Lys Ser Lys Lys Glu Lys Asp Lys
 20 25

<210> SEQ ID NO 24
 <211> LENGTH: 26
 <212> TYPE: PRT
 <213> ORGANISM: Mus musculus
 <220> FEATURE:

<400> SEQUENCE: 24

Glu Gly Lys Arg Lys Gly Asp Glu Val Asp G ly Thr Asp Glu Val Ala
 1 5 10 15

Lys Lys Lys Ser Arg Lys Glu Thr Asp Lys
 20 25

<210> SEQ ID NO 25
 <211> LENGTH: 26
 <212> TYPE: PRT
 <213> ORGANISM: Bos taurus
 <220> FEATURE:

<400> SEQUENCE: 25

Glu Gly Lys Arg Lys Gly Asp Glu Val Asp G ly Ile Asp Glu Val Thr
 1 5 10 15

Lys Lys Lys Ser Lys Lys Glu Lys Asp Lys
 20 25

<210> SEQ ID NO 26
 <211> LENGTH: 25
 <212> TYPE: PRT
 <213> ORGANISM: Galus galus
 <220> FEATURE:

<400> SEQUENCE: 26

Glu Gly Lys Arg Lys Gly Glu Glu Val Asp G ly Asn Val Val Ala Lys
 1 5 10 15

Lys Lys Ser Arg Lys Glu Lys Glu Lys
 20 25

<210> SEQ ID NO 27
 <211> LENGTH: 26
 <212> TYPE: PRT
 <213> ORGANISM: Xenopus laevis
 <220> FEATURE:

<400> SEQUENCE: 27

Glu Gly Lys Arg Lys Ala Asp Glu Val Asp G ly His Ser Ala Ala Thr
 1 5 10 15

Lys Lys Lys Ile Lys Lys Glu Lys Glu Lys
 20 25

<210> SEQ ID NO 28
 <211> LENGTH: 25
 <212> TYPE: PRT
 <213> ORGANISM: Drosophila melanogaster
 <220> FEATURE:

-continued

<400> SEQUENCE: 28

Glu Glu Leu Pro Asp Thr Lys Arg Ala Lys M et Glu Leu Ser Asp Thr
 1 5 10 15
 Asn Glu Glu Gly Glu Lys Lys Gln Arg
 20 25

<210> SEQ ID NO 29

<211> LENGTH: 31

<212> TYPE: PRT

<213> ORGANISM: Sarcophaga peregrina

<220> FEATURE:

<400> SEQUENCE: 29

Glu Gly Val Ser Ser Ala Lys Lys Ala Lys I le Glu Lys Ile Asp Glu
 1 5 10 15
 Glu Asp Ala Ala Ser Ile Lys Glu Leu Thr G lu Lys Ile Lys Lys
 20 25 30

<210> SEQ ID NO 30

<211> LENGTH: 28

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 30

gctgcggggtc tcgacgatga gtgcgggc 28

<210> SEQ ID NO 31

<211> LENGTH: 29

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 31

gcgtctagaa ttcacttggc tcctcaggc 29

<210> SEQ ID NO 32

<211> LENGTH: 38

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 32

ccggaattcg ggttttttgt taatgaaaat ttattaac 38

<210> SEQ ID NO 33

<211> LENGTH: 29

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 33

tcagagcaga tgaactcgag cagtcagg 29

<210> SEQ ID NO 34

<211> LENGTH: 61

<212> TYPE: DNA

<213> ORGANISM: Bos taurus

<220> FEATURE:

<400> SEQUENCE: 34

ccaatttgaa ggaggaattc ccgccccac catgaatgat gtgaatgcc a acgacctgg 60

-continued

a 61

<210> SEQ ID NO 35
 <211> LENGTH: 22
 <212> TYPE: DNA
 <213> ORGANISM: Artificial sequence
 <220> FEATURE: Kozak consensus sequence
 <223> OTHER INFORMATION: Synthesized by oligonucleotide synthesizer

<400> SEQUENCE: 35

gaattcccgc cgccaccatg aa 22

<210> SEQ ID NO 36
 <211> LENGTH: 674
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:

<400> SEQUENCE: 36

agaagaaaaa ggccaaggca tgctacaggt ggattttgca aatcgttttg t tggagggtgg 60
 tgtaaccagt gcaggacttg tgcaagaaga aatccgcttt ttaatcaatc c tgagttgat 120
 tatttcacg cttcttactg aggtgctgga tcacaatgaa tgtctaatta t cacaggtac 180
 tgagcagtag agtgaatata caggctatgc tgagacatat cgttggtccc g gagccacga 240
 agatgggagt gaaagggacg actgcgagcg gcgctgcaact gagatcggtg c catcgatgc 300
 tcttcacttc agacgctacc tcgatcagtt tgtgctgag aaaatgagac g cgagctgaa 360
 caaggcttac tgtggatttc tccgtctctg agtttcttca gagaatcttt c tgcagtgcc 420
 cacaggaaac tggggctgtg gtgcctttgg gggtagatgcc aggttaaaag c cttaataca 480
 gatattggca gctgctgcag ctgagcgaga tgtggtttat ttcacctttg g ggactcaga 540
 attgatgaga gacatttaca gcatgcacat tttccttact gaaaggaaac t cactgttg 600
 agatgtgtat aagctgttg tacgatacta caatgaagaa tgcagaaact g ttccacccc 660
 tggaccagac atca 674

<210> SEQ ID NO 37
 <211> LENGTH: 200
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:

<400> SEQUENCE: 37

aaaaatagtt gtcaagactc agaagcagat gaggagacaa gtccagggtt t gatgaacaa 60
 gaagatggta gttcctccca aacagcaaat aaaccttcaa ggtccaagc a agagacgct 120
 gacattgaat ttaggaaacg gtactctact aagggcggtg aagttagatt a catttccaa 180
 tttgaaggag gagagagtcg 200

<210> SEQ ID NO 38
 <211> LENGTH: 29793
 <212> TYPE: DNA
 <213> ORGANISM: Caenorhabditis elegans
 <220> FEATURE:

<400> SEQUENCE: 38

gatctcgaag taaaaactca cgcagaaaga gtcctcctc ctttagcatg a gaatccaac 60
 tttgtaatga taacactggc aacatcaaca gtttgagaga aagcacgtgc t tgggcttca 120
 caagcttgtc caatagaagc atocatcaca aaaacaacat tatctggtgt a actgcgttg 180

-continued

gaaacttgga gcatcttctc gaaaagtga gcttcttgct tgtgacgacc t gatgtatca	240
acaatgatga tttcgaaccc ttcttgctgc aaaacaaata ttattaaacc a tttttctgt	300
gataaattac cgtgaatctt tctactcctt cggcggcaat ttttacgggg t caatttcag	360
agtatgatcc atagaagga atacgagctt ttgtggcatt ttgctttaat t gatcaaaag	420
ctccagcacg gaatgtatcg gcacagatca gacatgtttt ccatcctttt c tttggtagt	480
aatacgccat ctgaacttga aaagtgttga aaagtgttg gaagtttact a attaaaaa	540
tataatgttt gatggtgtg gagctttcta ttgtaattca tggaacgaac c ttggtacia	600
gtcgtagttt taccggaacc ttgaagacca acaaacatga aaacgttgcg a cgtcctttt	660
gttggtgtga aaggagtac accaggatcc acaagcttca gcagttcatt g aatactgtc	720
ttctgaatgt accgacgttt gtttgctcct cgcagatct cttcgaaatt a atcgtcttt	780
ctgaaaatat ttattaaatt taaatcttaa atagcgtaaa aatctacttc a cgttgtcct	840
taagttgctt tacaagcga atatgaacat cagattcaat aagagctgta c agactctt	900
tcagatcaa atccagctcc ccctcattga taacggtgct ctgaccgagc t ttcgatcg	960
catttcggat tttccgcccc aaatcggcca aaaccatctt gaactgaaat t tgaatgct	1020
ttaatttgtt taagcataga attaaacgcg ttttaaatcg agagcaccat a aaaacagtt	1080
tggagaaaa tcgataattc ttgtaggaga ttcagtcctt gtggttttct t cggcttct	1140
aatcattttt tgacgacata gtggtatttc acaataggtt tttcaagac a caacagatt	1200
tttcacaaag agtagagaag aaatgaaaa ctgtagattt cttctcgaag a gccgagaaa	1260
ggcaaggtat tggaaagtta aaaaggaat gtttctttat tcttttttca a aacaataat	1320
aaatgaaaa tatatattta tagataacaa tttcagacag ttaaaatcac g tgaaaaatt	1380
caaatctcaa cacaaaaatt gacgagtga acccgttgt tgcgccttga a gagtaacgc	1440
ttgcgcgttt gacgatttta ttgacgcgtt totggtgcat gcgggaaatt t tttattttc	1500
aacttttttc ctgtttgttt atcctttttt aattgaattc tcatgattg a aagctttga	1560
aaaatattat tttgctcaaa aacatgcgtt ttgtaaaaca ttgattagat t caaggcaat	1620
taatggattt ttgacgttc caaaaaaag gaaattcatt tttgaaaat t ttgataatt	1680
taataatgaa aaatgttcca tagatttatt caatgccatc cttctctata a tctcgaact	1740
tccgcatcct tcaactgtgg tagaggtatt tgcaatacca tatagtcgta a taataaact	1800
ttagtgaaca aatccaagac atcagctctt gagtaaatga atgatttata a aaactgctg	1860
atcttctcgt aggaagaag agaatcagct aataatcgt cgttgtctat t ctgtcaggc	1920
cgcttaaatg ttaaaaaata aaaacgtttt aagctaattt tgtatgtcta g aaactctaa	1980
ctcacaagca tttctgcata cgcggatta gttggttttg caaaaagcga g taatctaca	2040
aaagtgaatt tttgattcat ctcttcatt tcacaaaacc aatgtgtg t acgtatttc	2100
atatgatctt catccacttt tttagttttt gaatgtattt gtgtgagttg t gtccagatt	2160
tgaataagat aacatctcag atccaacttg caattgaagc aagaacgac t tctctgaaa	2220
ttttatatga ccttaaaact tatacttttg tagtttcgct gatattctgat c gttcagttg	2280
tataggtatg tacatctcta ggtttatgtg ctacacgaaa atataatttg t tttacctaa	2340
cacacgcatc cataaaatga tctacaaatc gttcaattgg atcctgtctt g gaaataata	2400
atctccaatt cgtaaagtgt gattcaact cttttctcg tttcaaatcg t cgatatccg	2460
caaaatagt tagtgaatca ctatcacaca ctctgaaaag cacaaatctt a tttctgta	2520
gttaataatg aacctcacga ttcattcatta aatctctctt ggagcccgca t aatactgct	2580

-continued

tgccaatta	aagtatcagt	ttcacagatt	gcagttctat	catttccgat	a	gcotcaaat	2640
aagattta	cttaagcgag	tgttctgac	aatttaata	tttgatactc	a	ccgcaagtt	2700
tcttcgaa	ttgttcgaaa	gctggaat	tagaatatcc	ttcaaaactt	t	tttctcgc	2760
cctcatca	ccataataag	ttttgatcag	caatatattc	gaataaatta	g	tctctgata	2820
aatctcgt	cacaatcttt	ttttctactc	taaagaatac	aattttgata	a	gaatgataa	2880
taattata	tataatagtt	cgctcgtgag	ttgatgaaga	ccacataatt	a	gtttaatgg	2940
caagctat	gaactgttga	ataactaatag	gacttagcaa	atcttatctt	g	aaccttttt	3000
cattcgaa	aaaatgaga	tcgaatctcg	ttcaaaactgt	ggagtagtca	g	ttaagaaac	3060
ttgtttct	ttgtgagga	gacactggag	aacgtgaaag	tattacccat	a	cgcaatatt	3120
tttgcggc	aaaatagcgt	accgggtctc	gacacgacag	tttttaaaac	t	tgtaaatag	3180
gtatgtaa	gaaaacttta	attttaaacg	tggtgtttcg	gaattttcat	c	gttttgta	3240
tagttatt	acaataaatt	atttatgaaa	aaaaaactaa	aatataacta	t	aataacacc	3300
tgaattata	caaatcgatc	gaaaaaaaac	tatgaaaaaa	atggatgaaa	a	ttccgcagc	3360
aacgagagt	tgaatttca	gtattcttta	aaggcttacc	gatttcaata	a	atagtgaca	3420
ctgaaaatt	tagttttaa	actagtgtgt	tagtatcatc	aaatattcaa	t	ccttcaaaa	3480
attcctcaa	attaacgtat	tttctcta	tgtcttcatt	atctaaaaaa	a	agttgcaat	3540
atatttttc	aggcagaaat	agactttcac	aaaacacatc	gacacttcga	a	tgagcaaga	3600
agtttatcg	actgggtgat	cctgtcactc	aagacgttag	ttatagtttt	t	attacttga	3660
acattatcat	ctttttacag	gagaaagact	acgaagacta	tgctggagtt	g	gtttcgcgc	3720
atcaagtcc	gacaatgaaa	aggcggaagt	tgacagaaca	tggaataact	a	cagaatcaa	3780
aagaagatc	tgaagagcca	aaaagccgtg	acgtatttgt	ctcctcgcag	t	caagtgatg	3840
agagtcaag	agattcggct	gaaaatccgg	agatcgctaa	agaagtgtca	g	aaaattgtg	3900
aaaatctg	agaaactctc	aaaatttcta	atattgagag	tttgacaat	g	ttactgaaa	3960
gatctga	caactctgat	aatcacaaaa	gtactgaacc	aatggaagaa	g	atgtaaaaa	4020
acaagtcaa	tattgacgtt	gcgattaatt	ctgacgagga	tgatgaactt	g	ttctggaag	4080
agaataata	agaaatgag	gatggagaac	aagtacaaca	ggtcaggaaa	t	ttacaagt	4140
gaatgaa	agttaatcac	caaaatgaat	aaggacattt	cccacagaa	a	ggtctctcg	4200
aattttagg	gtaatgttaa	ttttttgctg	tagtttttcc	cattgtttga	a	atttttgcc	4260
aaaattag	attgcatacc	cttcatgttt	ttgaagattg	tttaggaatg	a	gaaaacatt	4320
ttggacgct	ttattattag	gacacaaaac	actttttg	gaaaaaacag	c	tcgtttaa	4380
aaaagcttt	tccaaaaaat	ctgacgcaag	gcttgtgaat	tttctgtttc	c	cctgatttt	4440
taaaattct	cctaaagt	tttgctaata	ttttctgcta	tcgcgtaatt	t	actagtga	4500
tcaaaaaa	atttttttt	tttcatagat	ttttataag	ttttgaaaa	c	atagattta	4560
aaacttaac	ttaaattttg	acaaggcgag	aggaaaaaat	taaaaattgc	t	gaacattca	4620
gatgccggt	acctattttt	tggttcaaaa	atcccaatat	tacgcgtctg	g	gttatagtc	4680
atttgcctt	attaaattaa	tggtgttctc	tggaaaagta	agttctgttt	t	gttttcagc	4740
ttatcactc	atcaaacgga	aggaaaggtt	gattaaggaa	agtaaacata	t	tttatgttg	4800
ttcttctc	ttctccatt	tcgcaataat	ataactcgag	aaatatagaa	t	tttgttoga	4860
agttttctt	ttccttcaac	attttaata	ttgttagtat	taccagaaa	a	atagaaaa	4920

-continued

atcgaagaaa	tttgcaaaaa	agcagacgta	gaggctacgt	acttcttaag	c	acgcccctt	4980
ttcttttaaa	ttgttcggt	cgtaccgaga	tccggtacct	tattttacia	c	gtttctgt	5040
tccaaaaata	ataatgtact	gcagttgtca	caggatttat	tcgctgatga	t	caagagcta	5100
attgaataatc	caggaattat	gaaagacact	acaactcaac	tgatataac	a	gattctgaa	5160
gtggagactg	ctcaaaaaat	ggaaatgatt	gaagaaactg	aagcagattc	g	acatttgta	5220
ggcgaggatt	caaaagttag	acaaaatcat	tctgacaagg	attcctgcca	g	cactcagtc	5280
aagagcgagt	cacggcaact	cggtcacaaa	ccatttctaa	ttagtaaaact	c	tcaaaaaacc	5340
acaactaaat	agcttaaaac	ctttgtaaat	tagcttattt	ttgctaatta	g	caatgattt	5400
taagctaatt	agttgtggtt	tttgagagtt	tactaattag	aaatggttt	g	gaccgagtt	5460
gccgtgactc	gctcttgact	gagcacaagc	aaacttttgt	ggatgttgag	a	atcagcggc	5520
aaagtggcac	tactagttag	gaagttgacg	cagatttctca	gattaatttg	g	taagacaaa	5580
gaaaataata	atthttattac	ccagatgcat	atthttcatga	ttctgatgca	a	aaaaatcgg	5640
tacccgatct	ggatactaca	atthttttaa	aatgcgaaaa	ggtttgcacc	t	ttaaaaaaga	5700
actgcaattt	caaacacttg	ttgctgtgga	ttgtttatcg	gtttttaata	t	tttttggtg	5760
agagtaaatg	agaaaagcga	gttcccgcat	tatctgtgtg	cgatttgga	t	acagtactt	5820
ttcaaagacg	cacaccattt	tgatataaac	aaacatttgt	cgtgctgaaa	c	cggtgaccg	5880
tgattttgca	ttaaaagtgt	caaaaattca	catagttttt	ataatttag	c	tacgaaaaac	5940
tgtgaggaca	tccagttcaa	gtttcctgtc	aactgtttca	acatgcgaag	c	ccctgcaaa	6000
aggacgagca	agaatgtatc	aaaaagagtt	ggaaaagcat	gtgattgcat	t	tactgaggg	6060
aaatctcaca	ctacaaccag	atthtgaaca	agttgatccc	gacagaaact	a	tcgatattg	6120
tacaattccg	aactttccag	cttcccaagg	aagtacgttg	ttcaataaaa	c	atactaggt	6180
atataattaa	ttatthtcaga	acttcogaaa	gataatcgat	atggcccaaa	a	atcgttttg	6240
cctcaaaagt	ggcgagaatt	tgattcggta	catttctatt	gaattaatta	t	atactactt	6300
actgaaaaa	ccatggagaa	agaatgcaaa	aaattgaatt	ttaaaaaacta	a	ttttttaat	6360
tttgctaaa	ttttcagttt	gaatttaatc	caaaatgaaa	actgcgacca	a	tcaatgact	6420
tttcaaaatc	acttttcaac	caatcaaacg	gagtggtctg	ggctcgaaga	c	gctgattgg	6480
ttcgaaaatg	ggcgtggtt	ctcatttttg	agggaattca	aaaaaaggca	t	ttggtcaca	6540
gttgaaaatc	atgttttcaa	aagatgcatt	ttttattcct	tctogatttt	t	tttgatttt	6600
cttttggtg	atthtctgaat	ttaaagggtg	tgtagtcgaa	tttttttatt	g	ctttattag	6660
actcaaaatt	ttctgaaaac	gccaaatttc	ataatgaaac	ttcttgaaaa	c	tcttcagca	6720
aaaagttagc	acggctcaaa	aaatggccta	aaattagtta	agattggaga	t	ttgaccgac	6780
ttgtcaatgt	cgcagcggct	ggaaacaatt	ttttttgaaa	tcaccgtcaa	a	ttttaagta	6840
tacaacttga	ttatthttg	ttttaaactt	tatttaggta	tttaaaagtc	g	atggacggc	6900
gagttttggc	tcaaaaaaat	taaaaatctc	gccgtccatc	gattttttaa	t	accttaatc	6960
aagaataaaa	caaaaggtag	gcaacttgta	tattcaaaat	ttgacggtga	t	tgcaacttt	7020
taactaattt	caggccattt	tttgagccgt	cataactttt	ttctaaaaag	t	tttcaagaa	7080
gtttcattat	gaaattcgg	gttttcagac	aattttgagt	ctaataagga	a	ataaaaaaa	7140
atctcagtag	accaccttta	agaaaattht	ggatttccgc	tacgctaata	c	acctttaat	7200
caaaaatatt	tgaagttatt	caaagttaaa	gaattatatt	ttcagagggg	c	cgtagaaga	7260
gactcatatt	tctatthtcaa	acgtaagctc	gatggatatt	tgaatgcta	c	aaaacaact	7320

-continued

ggatatttta	tgtttgtgt	aagtttttga	aatacaattc	gtttgaagat	t tactctatt	7380
ttcagggaact	tttgacacaac	atgtgggaat	ttgacccaga	catcacatat	a aactgccag	7440
cactggaaat	gtattacaaa	gagatgtcgg	aacttgttgg	tagagaagag	g ttttgaaa	7500
aatttgcaag	agttgcccg	atcgcaaaaa	ctgctgaaga	tattctgcca	g aggtatgat	7560
ttatgagata	tacagcattt	cctctaatag	tattgcatat	aaacatttca	c tttgaggtt	7620
atatcttgg	ttattttaaa	aatatcaata	aatacaaaac	aatagaaaa	t gataaaaa	7680
acattttgtc	agttgataat	ttgggtatag	tattcattca	taatttgatt	t ttttagcga	7740
atztatcgct	ttgttggta	cgtcgaatca	gctacctga	gccacaagca	a tgtgctgca	7800
cttgttgca	gaatgttttt	tgcccagccg	gacagtcctt	tcagtttctg	c cggtgagta	7860
atacaagaat	gctcatat	ttagaatcaa	tatttgcaag	gaactttaat	c ttacgtacg	7920
tcttaagatg	agcattttcg	cacatatctt	acgcgcacga	gtctcgacac	g cgaacatcg	7980
agctctctgta	actcgtatca	atttacaagc	cgttattaca	tcagttttta	a tgaatttta	8040
agaaaaatcgt	gcaaaagtag	tgctcgagagc	cattcgcgta	agatatggtg	a gatttatca	8100
tttttagaag	tctagtggat	atctaacaaa	actttataca	tttttatttc	a gaattctct	8160
cgctcgataa	atctatttgt	gtggagaaaac	ttaaattcct	gttcaacttat	t tcgacaaaa	8220
tgatcaatgga	tccaccggat	ggtgcgctca	gttttagact	tacaaaaatg	g ataaagata	8280
cgttcaacga	agagtggaaa	gataaaaaat	tacgttctct	tcctgaagtt	g aattccttg	8340
atgaaatgct	tattgaagac	acagctctct	gtacacaagt	tgattttgcg	a acgaacatc	8400
ttggtggcgg	agttttaaat	catgggtctg	ttcaggtagt	tatttaagg	a atataagaa	8460
tttgaagttt	tatttttttt	atgcaggag	agatccgttt	cttgatgtgt	c cagaaatga	8520
tggttggaa	gttgttgtgc	gagaaaaatga	aacaactgga	agcgatttca	a ttgttggag	8580
cttacgtttt	cagttcttat	actggttatg	gtaagtctag	actttcaaaa	a aaactgttc	8640
caatatgtca	atatatttca	ggtcatactc	taaaatgggc	agaacttcaa	c caaatcatt	8700
ctcgtcagaa	tacaaacgaa	tttcgagatc	gttttgacg	tcttcgggta	g aaactattg	8760
caatcgatgc	aattctgttc	aaaggatcaa	aattagattg	tcagacggag	c agttaaaca	8820
aagcaaatat	cattagggaa	atgaagaaag	catctatcgg	attcatgagc	c agggaccga	8880
aattcacaaa	tattccaatt	gttactggat	gggtgggatg	tgtagcattt	a atggggaca	8940
agccactgaa	gtgtatgtta	tttcatctgt	taaatattga	agatggagga	g agtgaatgg	9000
ggattttgct	tcttttgcaa	aatggcctcc	ctatgtacct	gaaaaaaaa	t gaaaaaatc	9060
gagaaatatt	gaaaaccaa	caacgaattt	ttcacaattt	tgcttaaat	t ttgaatttt	9120
cgccaaaatc	ggaatcagcg	attcgtctca	cccatttttc	cgccaatcat	t tataatgtg	9180
cggagctcaa	aaacactgat	tggttagaaa	gtggcgtag	cttcttattt	c ggaggaaat	9240
tcaaatagg	aagttaatct	aaatataaac	aatctcgta	aaaaatggtt	c ttttttcaa	9300
tcttccctat	ttgtttaaat	ttttcttttt	aaagatcgtc	taaaagctac	c agtatctga	9360
ttcaattatc	ggtttttttc	agtcataatc	caagtaattg	ctgccggagt	c gctgatcgt	9420
ccaattcatt	tctgttcatt	tgtagaaccc	gagcttgccg	caaagtgcaa	g aaaattata	9480
gaacgaatga	aacagaagga	cgtaacactt	ggtaagtcac	gtttttcaat	c ttcagttga	9540
tttgaaaaag	ttgtatcgag	ttggaacag	cttttaatct	aaattctgct	a acttacagg	9600
catgctattc	agtatgataa	acaacaccgg	cttgccacat	aagcactttg	a attctacgt	9660

-continued

cttcgataga	atctctactt	atctcagtag	ttcggagat	gttgagtctt	c	gaaatcatc	9720
accttcagta	tcccagcat	aattcgaatc	gcccacacgg	ccataaagac	c	ggttccttt	9780
cgattaaatt	ctgttaaata	tgcatgctcc	gtctttaaaa	aatcagtcct	c	gtattttaa	9840
acgttttgat	tttaatgttc	atattattat	ccgaaattag	tatactcgcc	g	tcatgaaag	9900
cccagatata	ctagtccgca	agtcagaaa	tttcggagc	atcgtcgtga	t	atatgaata	9960
aatacattcc	tgtttttcac	aagtgtagt	tgaaccaat	ccatgcagac	g	tttatttct	10020
gaattaattt	tgaacagat	ttcagagaca	gtgaggttga	cattagatat	g	ggcaagtaa	10080
caataacagc	agggcagtta	ttatgattat	ggatgctgat	ataggaaagt	c	agaacagta	10140
taatcgagca	gaataaaaag	agatgagaag	ataggcgaga	ataaagaacg	t	taacgaaa	10200
tcactgaaga	gctacatttc	caacagaata	agaaatgtag	ttggaaatcc	c	taatcaaac	10260
agaaaagcga	gaaatcatga	ctttcgagat	aaagagattt	atctgcaaac	a	attcttgaa	10320
cataaaatta	aagcaccaca	gactgtccaa	attataaaat	cagtttctcg	c	tacagtctg	10380
gggttactct	agttccattc	aaaaacttct	tgcaacaaa	gagaaataaa	c	agaottgta	10440
cgggacacat	ataaaatcta	agcatgcttt	gaaaagcga	gaacatacga	t	ctattcggg	10500
gatacacata	tatatatata	tatttcatct	catctagagg	atcaccatcg	t	tactcatca	10560
aattggttgg	tgtggtggaa	gttatgaaaa	gagcaatttt	aaccgaaaat	c	accaaaaa	10620
gaaaccaaat	taatgtataa	tcgacagaaa	tcatgatgag	atgatgattt	g	ctctagca	10680
gaagtttaga	agcacatgct	atcattcatg	ctcacgatga	cgataggttc	g	ttatgcatt	10740
cttgaagcca	atgacacttc	cattgctcct	tctcttgcgc	tcacacaatt	t	ccattctcg	10800
tcgtaaatcg	ttcgactttc	gaataaccac	atcttaccgg	gcggcacttt	t	ggcctggtt	10860
tggcagatct	gaaataaaat	cttttcataa	tttaaagtct	gatatcccga	g	aaacaatag	10920
ctgaattgaa	acagaagaat	aaatctcacc	tgaatatcca	cgtttgcaag	t	gaagtttct	10980
ttcacaaaag	gctgtgtttc	ctcatagaga	tcacttgaac	cgtaggagaa	t	ggacggagc	11040
aagtcttcca	gcttctcat	gtgctgttca	tttgccggtc	tcacgtgctt	c	tttacggga	11100
gcagtggcgg	cagcatacgc	tgatttagat	gctcgcatag	tgcoattcct	a	cgactttga	11160
ggcttccttt	cattgttcta	atgctgttgc	atgtgaaacg	gtgctgtttg	a	aatcctcgc	11220
atgacttgag	acgattgatg	atgctgatgt	ggaggaatgt	ctgcaagatt	a	tagttatgt	11280
atataaaatc	aaaaatttgt	gtggttccca	ttttaaata	aaaaaaaaat	a	tttttacgc	11340
actttgtcta	ggcaaccgat	aactatttcc	tcgctggcga	ctacttctct	t	attgtgagc	11400
attatagctc	atgttctcat	gattgagttg	acctgaacga	tcaaggtaa	a	actaggcct	11460
aaaactagtc	aaaattactg	agtttctcct	tccacgtcgt	ctgtcgagca	g	gctccgagt	11520
acatttttac	tggaaaacta	taataaatta	caaaaatcac	gccgaaaatg	g	ggaaaagaa	11580
ttgaaaaatt	gaaggaacac	agaacatttt	ttcaatgcgt	ctctcacgtt	c	gagactact	11640
gtattctgtg	tgagacccaa	ctccctcata	aaagcatgcg	cctttagttt	t	ttaatttaa	11700
ttcatgttgc	caatatttgc	caattaattt	caagagactc	tgattgaaag	t	gttataatt	11760
aaactacata	tatttaagct	ttcagcattt	ttttcaatgc	acttgagacg	c	aaattgaat	11820
aatcaggcac	gtaatgtggt	ttcagggagc	actataaatt	gtacctttgc	t	atccagtgg	11880
gttctttaa	tttccattc	caatcgattt	tttctccac	tctggcagtt	t	ctttgtcat	11940
cactggacga	gggcatttga	atgggagatg	attcatgtga	caatccacac	a	tcctgcaat	12000
aatgacatta	tttttaaaa	atgttaagat	gatatgctta	ccaggagtaa	a	tatcatatc	12060

-continued

cttttcttta ttagttggct tagccttgcg gccacgttta ccatttgaca t tatagttac 12120
ctgaaaattc aaaaaattag atattcaaaa aggtataaat ataataataa t gcgatttgg 12180
taaatacggg tgtaatgggc aaccattct atacagaaa accaaaaaat t cccgcaaaa 12240
ttatTTTTTT ccgaataaaa tgatctactt tgTTTTatgg tgccgctcta t gtgttatga 12300
cccttcgatt agtagataga aaagaaaaag gaatgtacga gaatatcgtt t attatttat 12360
tatttgaaaa atcccagaga cataaaaaat cacacagaaa agggaaacag t atttctgac 12420
aatgttcaaa agtttggttt caatcagcac taataatgtg aaaggtaac g tatcaatag 12480
tgatattttc ttattaaaaa actgttcgag actacaagaa ggctgaaaa a gcccgcaac 12540
gacgactaaa ttcgaaattt cgaattaggt ttaaagatc agaagatcgg c agaaaagta 12600
tctgataaaa atataagaaa tcggaatagg aatgcgatga ggaggtagaa a tatggtgaa 12660
gagatacaga agaatagggt aagatcggat gaacttgaag cactttttga g atttttgat 12720
ggtgaagtgt gtggatgtag acgtttcatg gaacatctga aaattaagat t tttctaaaa 12780
cacattttct atagaatata atagaatgcc aaatagagaa actagactta c ttgaatttc 12840
tttcgatttc tgtctttcaa ctttctaact gaaatcaact ttcgacgtgt t ctcggtgtt 12900
tcaacaacac catcaacaga aactcagca ccaattcag catcggaatc a tcagaagaa 12960
gactcatcgg aatccaaata gaaattggat ttagtattca tcaattcaaa a gaatccaat 13020
gatactgtcg attcagcaag ttggactgaa cttgatggtt gactacgaac c cattgaggg 13080
cgtcagggca gaagtcgaga gtatgaggat gcaacgtgga ttgatgatga c gtcaacaat 13140
ctttggtgtt gagatgaaga agtggctgat gcagatgttg acagacggaa t ggagatgag 13200
tgaagagcaa gaagacatct gaaaatttga aacgttgttt atgtggacag t actgtaaag 13260
atcttaccct ggatcataac tacttgccct ctgttttctc ttctcttgac t tctacttaa 13320
aagcatttcc gtctcgattc tccggttact tgaaaatcca actccagaat t ttcagcaca 13380
aagctgctct cccgaacctg agactgttgc accacgttga ggggttgaca a ggatctgaa 13440
atcagatgtt taaagcatg caagtagagc aacaatgta accaaaattt c tgaaacttt 13500
ttcgaatata gtcaaaaatt gacaataact cagtttcacc tatcatagtt t tggaaagtca 13560
accaaaatt tttgaaattt cataaaaatt ccaactttc taaaaattt g aagattgat 13620
atgattgata tgaaagtatt tatatatttt ttaacctggc agacgatact t caccattaa 13680
agacacacat gtggagaaga attattttac ttttagtaat ccaacgtttg c acttacctt 13740
ggagcatgca agcttttagt cattaaagct ggaattctag atggagtctt t cttggtgtc 13800
gacattgttg aaataaacat tcgtggttca ttgattgatg atgacgtcat a gaaccacgc 13860
ccagatgaca atggattacg gtagtcatca gaatcagtag attcattcaa t tttctagtc 13920
atctctctg ttttctgaa aattaaattt taattaaaga tctaacaaaa a tctggcact 13980
tacattaata agataatcaa catattctaa ctcatcctc gtttcattat t ttctaattc 14040
tggtctcttc tcatcgaacc gttcgggtgc attgtgtcgt tgcgggcttg a cctttttt 14100
gaatttctga aatgtttttc atgcaatttt tgttcttatt tgtgtgtcat a tacagtga 14160
aatcaaaaa tagtacaac taattccgtt tagtaaataa aaaatcagat t aaaatctca 14220
gcaaagccaa gatcttggcg ggtccttata tccaagtttt gttgccattt t atttcagat 14280
attcttttgc aaagtcagaa aatttgaatt tagaatcгаа tggaccatt t cttgtttt 14340
ttttgttga ttttttaact gtactttttt cgtcagcata ttttttact a ttaaaacag 14400

-continued

```

aatattcatg acaataattc cacaaaaaaa cgtactttta tatcatagtc g attggttca 14460
gaattggaac gagaaccttc gacgcgctga ttgtcagatt ctcgattgat g gacgacgtg 14520
ctgactgaaa atttctggat tgaaaaata ttcaaatgaa aaaataaatg a gaaactcaa 14580
agtctaaaaa atgaatgttg ttaataacga atatttctga tgagaagagg a tagagaaaa 14640
aaaacgagtc taataaaatg catgtgatat cctgcataaa aatcccttct t ttttcacta 14700
atccttgcgt caattcattc aatagaact ttgatttcta ttagagtga g gttgtttga 14760
acaattttta taaattaaca ataagccata aaacctcga acgtaccatc a tcattgagt 14820
ttgaaaaagt ggacgcatcc gagtcagtc cctctggaac aaatcgttcc a gagcactga 14880
aaacgacaac gttctcccca cagaatcgga ttgtctctc ggaattgtc g cctcgaca 14940
acgatcctga acctgaaaaa tttcgatttt tgtaagctca atggatttta a actgaaaat 15000
gtagtcaaga agtcaagaaa aactgatgga gttctaaatt cgggttagg t tttgaaaag 15060
atcgtcaaac aaacaaatgc ataaaactag gtagggaaca aatagtgaa t agaaaaatg 15120
aaagcgaca actgccggga gcaagagtac acacaaagaa aaaagttgc g gaagagcac 15180
agagagcgtc agtccatcag aactgcatag ataaatagat aaagagaaac a tgaacata 15240
aggccaccgg ggagagacga caggccagtt ttccgggtgaa gatgagagtg c gagaattag 15300
ataagaaaaa ggaattgtg atgaaacttt ttcaatcaa acttctagaa t tataagaga 15360
cacctaaagt aatagataa gtgttttaag tgatatttta gattcactgt a tcatgttta 15420
aaaaagatat ttcaaaaaa tatacctgat ataggaggcc tcctctgagc a ccgaattgt 15480
tctcgagctg tttccacgag catccgctca cacattgaca taggccgtcg a cagccagga 15540
gttgccacct gaataaataa ttattcaatt taaacctaat ttagtaatg t aactttgta 15600
aatgatggtg gatagctcat ataaaattg aattggttct aaagttatac a aattttaat 15660
ttcggtcaaa cttatgaact gtacttttga gttatactat tacaataata t taccocaaat 15720
tattgtattc agatttttgt aatcagtact aacagatttt aggcaacgtc c tgccagaaa 15780
catgggaata tatttgagca gtttttagta agttgccaca gcttgataa g ggaattgta 15840
tcaaaatgta cttaataact tctaagcact gacatagtga actacaaaag t cggatttat 15900
acaatgccac tacaaataaa aatattcaga attcgactga aaaatgagaa a aggaacaac 15960
tgaattggac acacgatgtc gtgattttca agaacacaaa aaaagaaaa a gaaatcgaa 16020
aatgttgttt gcctcttttt ctttttctat atgagctaga atctcgaatg c atgocctaat 16080
ggagccactc gctctcgttt ctctaagtct cttctcacca gtcttttgtc c aaaaattgc 16140
gatgtcgcag gcgtcccgtt tccgccgctc acggagacac cactatcggc a ccgatgat 16200
cgtgtaaaga caccgctctc gttgacttca attgctgaaa taagaggaat t agttttgaa 16260
ttgaaatct gattaataa aagtccccta ttcaatctaa ttaattttta a acacaaaac 16320
ttactattgt ttaatggtg tgacgaatta gaagaagttg attgaaatgc g ttaacattc 16380
cattcaaagc tatcttcatt ccattctgct cgttctttta tcctttcact c acgtctoga 16440
gggatgaagt tttcaacaat aagaagcctg aaaacttata ttattctatt a aaaaaatg 16500
aacatcaaat cctaacgaaa gaaaattctt ctggggggaa aaggagagaa t tgtgagaat 16560
aaagaacctg cgctgtcgtt atcaaatc actatttgaa ttcaaatag a atacgaaag 16620
aaagtgaag aaatgaaat gagtgagaat ctattaaatt gtaattgaga t atcactgaa 16680
cttacttcaa cttcaattcc ttgctcactt cogcaattgt ctgatccaga t cttgtcgtt 16740
cgtcggaaata tgctccagac acatcacgaa tctcattgcg agcctgaaac a ttcacaaaac 16800

```


-continued

cttatcttga cacctggtac atctgaagtc aaacctgtct caactttatc a acatcttct 16860
 ttaacttttt tgtctttgcc tccacctctg tccttaaate cgaaaatggt t gtttcaaat 16920
 ccacagtatc ctcttcctgc cgttccaatg cttcaacatc ttctctttcc c ttcgttttt 16980
 gttcggcaag ttctcttctc ttcttttcca gttgtgcatg ttgttctttt g tccttgatt 17040
 ctaaccttcc atctcttcca gatcctacga taagtcgact ttgaatggtt g ctattcttt 17100
 cagcaaccgg tgcttgttct atccgttctt tttctaacag atcatgcttc t cttgaattt 17160
 ctcttattaa tcgactcttt tcatgtttta tcagtgaatc atctttttga a ttgcttcaa 17220
 tatcatcttc aagttttgct cgttcagcat cataaaaact ttgagttgct c catccctcg 17280
 atcttgtctt tctctgtctc agttgtctac ggagcatttc aatttcttct t ggaattctc 17340
 gcagtaaagc atccttagga tcttcattaa ttttcggttg attcttgatg t ttttagctc 17400
 gatttgcata tcgtaatgta ccaagtgtct cctcaaaatt gtaacttgca g gtccaatac 17460
 aagcaaccat aactgtcttt gaattccac cgagagaatc ttgaagaagt c gagtcagtt 17520
 ttgaatctcg ataaggaata tgggcagatt tcgcatccac caatgcactg a ttacatttc 17580
 caagagccga taatgaaaga ttgattttcg tagcttcttt aaatctttcg c cagttgctc 17640
 ctgttttcga ttgccgttct gaaccagcta aatctacaag atttagtcca c caactgtaa 17700
 tatgactttc tccgtcttca ccaattcggg aacattcaac agtaatgata a agatagcgt 17760
 gggaacgaga cgaatgtcca ttcattgttg ttcgctgaa aatttttagta a aatcaaatc 17820
 caacggcgac cacagaaata acttacccta cagaccgatg cccatttctc c gaatcatca 17880
 ctctgtgat ttcacctact gtctctgta attttgactg aaactttgaa a tttatagtc 17940
 gtcttctatt tcagaaaact atcacttacc gttaaatctt tcacataaac t cctccatct 18000
 ggacgttctt taatttctaa tttcttattc gattcggctt ctaataaatc t cgaagtcc 18060
 tcctaaagat ttcatttttg taaatcacac atcctaacgc cttacctgat a aatttccaa 18120
 atagctagot ctaactaaat actcttgatt atgtgatgct gccatgtgct c aaaaatag 18180
 gtcaatacac ttatagatga cacctcgttg ttctggatcc gatgattttc c ttccattgt 18240
 gtgagtcttt ccagttccag ttgaccata tgcaaaaatc gtggcattat a tccgtttag 18300
 aaccgaatca actagatctc gaaaggtttc ttcataataga tccgattgtg t ggaactata 18360
 aaatatatat ttttaaaaaa gagaactcat aaaatcataa acataaaatt g tggagaaat 18420
 aattttgaaa aataactaata tttctatagc aggtgaaaaa aaagtgatgt a ctctagaa 18480
 ataaataatc ttacttttca tcataaattg catcgaatgt aaaatccttc g atggctcat 18540
 cttgtctctt tggatttttc agctcaattt gccaccgttg tggtcgcata t gtactattc 18600
 tgaaaatgat cgaatattca aaatataaat atttcaaact ttacttactt t gaataatta 18660
 tttgcaattt cttgtgaaga taacggtcga catctcacia ttacctgtaa a cataaataa 18720
 atacatttat ttgaatttgg aaatgataaa aactggatta tgaattttt a agctggtgg 18780
 tttttgatg agaagtaacg aaaaaagta caatttactt agagtcttgt g atttttctt 18840
 tcaaatgcaa aactcaactg aatcataaat agtgatgctt cgaaaagttt t tagaggaaa 18900
 attgtatttt tagtaaaaac taatatacgt tttggactta aaaaaaatt a tgttaaac 18960
 ttgaaaaatt acgtttatta gtgcttata taaaaacgg tttcaaatta a tttaaaatt 19020
 aaaaataact accttttggc caaaatcaga catttttaga actagcatgt a ctttattac 19080
 gttgaatata acttatgttg gaaaatgaa aatttgaaga caggtgaatt t tagtttttt 19140

-continued

ttctttttog	tacttctaaa	aaatacttca	tttatTTTTac	atTTTgagaa	c taatttttg	19200
aacatgtttc	gaacaaaaaa	aaagatTTTg	aaaaccccaa	aaaaacttac	t ttgacagtc	19260
tcctgttttg	aagatTTTTt	cattatTTTc	accatTTTTt	gtcactaaat	a tttggctat	19320
caatgtaggt	gtcaaggaaa	atTTTgtTgc	attcctgatt	tagtgagagt	g gtctggaac	19380
ttaagaagat	tagtttaatg	tggaaaaata	atcatattgt	atcgagaaac	g gaattttga	19440
agcaataacc	gctagagaaa	gtgactaaaa	accagaaatt	gtagtcgtTg	a atttcaata	19500
TTTTTgTtt	tatgtcacat	ctggacaatc	ggaaaaatat	gcatacattt	g aaattttta	19560
gaaatatttt	gaattaactt	taaaggaaaa	aatgcatta	aaaagattga	a aacatcatt	19620
gacgttgaaa	aatggagaaa	atTTTctaatt	tctcatcaaa	atattaaaat	a ttaaagttc	19680
ttcaataata	tgaaaaatgt	aataaaatgt	ctaaataagc	aaaaaaaaaca	g atcctattc	19740
attataaaat	gttcacacaa	gtgttacatt	tcgtacaaag	aagtactaaa	a cggatggac	19800
taaagtaata	ttgtcactcc	cgaaaagacg	aggaagaagt	aatcggaaga	a gatgtcTgga	19860
agatgagTga	tagtaaaaaat	acgaagagac	gcagatagag	agtttgagag	a aggagactt	19920
ctggaggaat	aaaagTggt	ttcaagatTg	gggacagaga	gggagagTgt	t aaaagagca	19980
caaaatgtgc	ataatatcga	tcctgcgcag	ttgagagacg	cagacaatgt	g aagaatTgga	20040
gcataTgttt	ctagtgaaca	ctcagaagta	gttTgtcatg	Tgtccgaaac	t ttggaaaca	20100
tatacatTTt	aaactTgacg	TTTTTgaatt	ataaaggat	ggagTgctt	c aaaaagtaa	20160
tcatagacat	gtgtagattt	taaattaaaa	cacaactaga	cataggatga	a tcagaagct	20220
taccataaca	ttgttgattt	atTTTaaaaat	gagaaaaagt	aaaattccg	a tagtcttct	20280
ttgaaaaaat	tcacagagaa	gttataatgt	ttgatgat	tcactgattt	g taatacatt	20340
attagtagca	Tggcttctat	gtatatagac	tatTTTtata	tcacatacat	g aaaaagTgt	20400
taaggcatgc	gccagggcct	gaaaacgc	ctacctaaca	ggggagctct	a gctottagt	20460
tattaatTca	agagactTTt	gaaactTgat	TTTTTgagat	tttattcaat	g attgTttta	20520
aaaaaaaaaga	ttatTTTgcaa	aaattacaaa	TTTTaatgtc	tatactctga	c atcggTttta	20580
gaacaatttt	agacagTctg	caatgaaagc	aatgaaataa	aatttcctTg	a aattataat	20640
agagaatcag	taaaatgtTg	cagattatTTt	gaaaatgc	gcaagaattc	g cagaaaaatt	20700
cagTgaagca	gaaaagTgcg	acaggagacc	gaagtctaaa	aaagtgaatt	a tgaataaaa	20760
acaaatcatg	Tgactggata	taattgaag	tcttgattc	gaaaagataa	t Tggagctct	20820
ttgcttaggc	caggctctag	atattttatt	gaagctTTTc	agaaatTgtc	a aaattatca	20880
ggaacagTtc	tctttgcaact	ttctctatTg	ctcaactacc	agggctTTTc	c tttttcttc	20940
aaaaagtaga	atTTTaaatt	ataatTTTaa	aattTaaata	ccaagcaaaa	a atcataatc	21000
tcatacatatc	atgtgatcat	atcataataat	catatagTgc	tcgttctTTt	t tttttcaaa	21060
aaattaaaaa	tttactagaa	ccaagcatat	gacaataaaa	tatTTTgaaat	t cactTTaat	21120
gggaaaaaaa	caagaaaatt	tcattaacat	tattgaaaac	atcgtTgTgca	a taggaaatgt	21180
agaaaaatcaa	atcaaaaatca	agtgagatta	ggaaagaatc	gaaattagTg	a gaattgTgaa	21240
aatctcgatt	TTTTaagTg	gattcttaca	cgatTTTtTc	gggatatttt	t catTTttat	21300
TTTgtagtat	ttcagcctag	acggctgaga	attctTTTca	aaccttccaa	t ttcaaagag	21360
attcttccat	aattTaatat	aattttcatt	cgatattagc	atccattata	t acgtatgat	21420
tcccctTTta	aaatcgattc	tcotTTTcaa	ctgactcatc	actTaaagaat	t gttgagTca	21480
tcaactgata	gtgagcagac	accaacaacc	atctcttag	tttccgtTtc	g tttattTTta	21540

-continued

ttttggaaac taacatattc aagaaaatta acttgaaatt agaataaatg t ttcttgcta 21600
 gatttttttg tcataagtat ttcttatttg gattataatt ttcatctcga a atcgtagag 21660
 agttttttcac tttttttttt tgagttctaa acaacttcctt cctcatcgat g atgaagttt 21720
 ttgacaaatc aactagtttt ttactcatat ctacatcaa tctatgattt t attcaaaaa 21780
 cagttaaatt tttttaacga aattaaaatg gtcacatcgac cgagcaaaaag c tttcagaat 21840
 caactgcttc tttaaattct ttaaaattca atcaactttt cgtgtccaaa g tcacaaact 21900
 acacctttca aaaaatattt ctacattatt tgcccacatc ttggcacagt t ttcttgcca 21960
 ttcttcaata ttttcttctc tgcgtttccc aactcttat tttctgactg t tgaactttc 22020
 cattgtatag actcaatttt actttcgttt tttcaatttt tttttctgcg a agttcggtg 22080
 ttaaacctoc attttgcaat attaaaaatt tcaatattgc cgtttttggc t tgaatctat 22140
 taaaattatg ctgttttttt ttcagaaaag accaaaacat gccagatgat a ttccaaaat 22200
 tgccacgaca cagaggaaaag aaaaatcagc cgaaaggttt gaaaaattta g aaaaatctg 22260
 aagttacttt tttaatctct tagacacacc ttggaaaaca caaaaactgc c tgctttacg 22320
 gcctcattat aacataactt cagcaattcc agttactctg ataacaggag t agccacgtt 22380
 ggcaatggga attgctcttt atttcggaca taatggatgt gagtttttag a gtttattat 22440
 cccaaaaaca aaaaatcaaa ttactctttc ctggtaataa gtaagaaaaa g ctaaagaaa 22500
 acaaatttct tgcataaat ttacattgta aaccgatagc acaaaaaaac a agtgcata 22560
 aaaactgtaa gaaaatcgat aattttgcta caatttcaca aagctaaaaa a tttttttta 22620
 ttttaccgtg ttagtaccgg aatgttctgc acttgagcct tactattagt t acacaaaat 22680
 ggatcaattt tgagcaattt gttgtgaatc tgacaataat tagtcctatt g atatagctt 22740
 taggccactc attcgtgttc gtaattttcg ttttcttga acttgtaaag g tacagtttt 22800
 tgaaaacagg gatgtagtcc aagtagtcaa atattgattc ttgtagcatt a gaacaagag 22860
 attgtgtaca cggatgtgac tctttcaaat ggaacacaag ctacacgaat t atgagaact 22920
 gaaatgggaa atcaaacatt taaatgtgca tatacaatta ctttgaatga c gattatact 22980
 gtaagttgag ttttaatttt taaatcatca aagaaaacat atgtatattt t tgcgaagga 23040
 aattttggat ctggtcttag gatgaaacga cattgtaaca ttttgattaa a gagccotta 23100
 gttggaagtg agtgtatctg gtaaaaacac aattcgaaaa tatttaacca a atatgtata 23160
 aagcctaggt tgaacctgct ctgcagttcc taatttttca cattattttt c ttcaaaaata 23220
 ttactatgat atttcaaagc cgggggtacc atcttaaat catcatttgc a agtatcaca 23280
 attaatgttc aacattacag ggcaagtgta agttttatta cggcttttcc a agttctatc 23340
 aaaacaatcg attatacttc aactcacgaa acgatcaaca gctacgtgga a aagttactg 23400
 aaactgacgg atgtgatcca ttagaatatg tggatgtaa tggaactaaa g ttccattg 23460
 cgccgtgtgg gaaagtggtc gattcaatgt ttaacggtgc atttcaattg a ttgcttaat 23520
 ttcagtattg caacattttt cattttttat aatacatcta acttcaaaaat t tgttttttt 23580
 ttcagatacc ttcgaattat tttatatcaa tgataaagcc tcaaacgagg t aacacgggt 23640
 tccatggaca actcgtggag tactcgtgtc aactgaaatg aaaagaaaat t cagaaatcc 23700
 gattcagag gaaaaccaga cattatgtga tgtgtttgcg gttgaaatga a ataagaaaa 23760
 aaaaataaatt aaactccatc ttttagggaa caatgcctcc gccatcatg a gatatccga 23820
 tctgtcaatt gggactaaac agtattgatc cagatgttgg cattggtttc g agaacattg 23880

-continued

atthttatggt ttggatgaag gttgcagctc ttccaaaatt cagaaaactg t atagaatac 23940
tgaatcgaca agttgatatg ttcagtaatg gattacctaa aggacaatat c agttgacca 24000
ttaattacag tatgtttatg ttaatgttga atthttatgtat ttatgcaaaa a atthttactgc 24060
aaaagtacac aataattcca cccaaacctg cttaaatatg gagatgcaag t tthtttgttt 24120
cagataaaca gtggctccaa aaaaccaatc tttgttataa aacctcacia a aatthttctcg 24180
atthtttcttt atthttggtc aaactthttga gaaaaaaggg aaatthtagaa a atthtttca 24240
agcgaaattg tcaaaatthtt tcaaaaccaa atthttgatttt ccagattttat t tthttgtcga 24300
cttgacaata gtaaagaaaa aacaagttga atthtttctat atgaattctt a tagctgaac 24360
atthtttgatc aatthtgaaaa taatcaatag acaatthtttc tccatactac t gattthttcag 24420
actatccagt ggatagtatg tctggcgaca agtacttctg tatagccaat g aaaactggg 24480
ttggaccag gaatctgttt ctaccagtaa tctatthttggt tgttggaaca t tcttacttc 24540
tctgtactat tctcttcata ttgatttggt taaaaacagag actgtctgagg g tctatccaa 24600
catgaattgg aaaaactaat tgaaaataga cggatgaact tcaaatthttg t tacaagagt 24660
tgaagtctca aaataagctg gtagcatgta ttgtacggga acagattthtt a tactthttgct 24720
ttgtaataaa aataaaatgt tattatatta gtctgtaatt ttatgtatag t tcaatthta 24780
ttgaaataca taataccctc ttcagtttat caatthaaagc tccaactatc a tctgctggt 24840
tgagattaat tctcagatga gggcatctga aatgtaaat taaaattaca a aataaataa 24900
ttgtaagtgc tatcagatat aacaaatgat catttaatta aggaggaaaa a caaaacatt 24960
aatthtaaaaa atthttacaa aaacaaaaaa aaacggctcaa atthttthttc a acaaaacaa 25020
agtaaaagcta atthttctatta aagttgatct aaactactggt tgtgtaggca t actatagtt 25080
gattthcaacg ggaagaaggc caaatcagca agtgtacatt gtgttctgaa a aattgaaat 25140
tcaacagttg aatataagta gaaactctac ctattgctaa cattthattgc a atthttctg 25200
tgthtttgaac aatctcgaga tctgctccatc catcgataa tctcgatga t ttgatgaca 25260
tctcatctac agcttccaat tctccaatta tctgatcctt cagthttcagc t taatatcaa 25320
acgattthttg aatthttctg atatthttgct cataaactct cgagattthca g atthtaactt 25380
gttgaatctc tthtaattgc tctgaaatct tcttctcaag aactctthttc t gataatata 25440
gctcagtgat tctctgagtt ctthttctca taatcatttc agtattcatt t gthttctctt 25500
taacattthtt tcaagthttt tcaatctthtt gtacataatc acagaaatga t caaccacct 25560
gcaaccattg cgggtcattt cgcatagttt tgagtcctcc gggthtttca a gaattgcca 25620
caagactthtt tctctgttca agthttcagtt tthttcaatgt ttgctcaagt g ggaattctc 25680
tgacctcctg gthttgcctct tctggaaggat cagaaatthtt ttcagaatgg a acgtcaaaa 25740
thttctgttc caaatgggga tatgttctac tagtccctg actagaagtc t cactagaga 25800
thtttaagtgt cagthttctga acattacgct ccaatthttcg agaattthcca g atthttcagtc 25860
tcacatgact ctthtaatttg acaatctcat cattcttatc aaagattthga t thttcagtg 25920
cagagacctt cacttgaca tctthttgtct tccatgaca tccatctgcc a acatcttga 25980
tcatgttctc atthcgaatca atctgthtttt tcatatctthtt tatcatgtct t catcagthtt 26040
tgattgtaac atthttgctthtt gaaatthttc gthttcacaga atthc aaagca a thttcaag 26100
aattgttgaa atthttcgagc atcaaaactthtt caacacctc ggattthttatca t tctthttgcaa 26160
cattccgac atthttatgt tctagthttgata tagaattgggt atcatcaact g agaatactg 26220
ttaaattctg gtgttgaaat tccaaaaatt tctccatgac thttthcaata c tctctctct 26280

-continued

gataaacatt gaaggattct tgaagcattt taatctcttc attcttttct t tgagttcgg 26340
aaaacagaat attcttttct ttttggaatt ccaactgtcac taagagataa t catctttgt 26400
tttccaaaga tgtaaccagc tgcgatggat tcaaactttc caagtttttc g taagatagt 26460
caacttcgct ttcaagtttt tgaatgatct gtttggttgc ttccatcttt t ttgttataa 26520
gtatagatc cttgaaataa agagaaaacg tgactatgat cttgtcaata g tttccagca 26580
agtgaaaaat gtcataataa tctccatcat ttaatacttt taatttgtcc a aaagttgat 26640
ccatcaactt ctctgttttc ggtgtttcat caccaattaa tattccaccg t aacgttaa 26700
ccggatagtg cgacaaatca taaaattggt tttgaagatc ctcatattta g ttttgagaa 26760
ttgagaaatc ttcaatttct attatcaaca acttcttcaa tcttctagct t ctgtctctg 26820
cttttatttg gaaaatttgc aattgcattt cagctagtat tatctcatct t cttgttcat 26880
cgattaattt ctcaacttct tcttttaatt ttttgatatt ttgatctttt t ctgcatag 26940
cattttgaaa atcttctagt gttgcaact gatattcagt ggatgtactc g tggattctt 27000
gagtcgaaat ttcagtgctc ttgtttatct gttttacttc ggagcttctg a ctattttct 27060
ctggagtcca aaactgtctc acttccaaaa aatgtgtttt tttgcttctg a aaaacatat 27120
attaagtaac atctttaaga tattcagtg cacttacatt tttgaaatat t tggtgacaa 27180
actttgaatt atcaatctga attcttctgc ggttccagtg aagcaagcat a attctgaaa 27240
ataaaaatta cagcttttga aaccaatgaa acgaaacaac tattgtattt a aaaaatgct 27300
cacttcaact ccatttctct ccaccgctgc ttcttttttc acacttttcc a gtttatcaa 27360
ttaaaaattc aagtttctgt tgttcaggtg aaggctgaga tgctgtgaaac g acatagttc 27420
tgaaaatagt taatttaaat gtagcagaaa aatcttttct agaaagtaaa a aaaatcagt 27480
aaaaacaagt actaagagaa attgaataaa ccaatcacia taatgacttc t taacaagct 27540
gaaaaataat gcaatagcaa agaaaaacga gtagtttcgg taactccata g tacattatt 27600
tcgttatttg gatcatcata tcatttattg atgaggatat tatgagttaa t tctaataac 27660
ccgagagtaa aggcacaaaa tagcatggag tgaaaaaacg gatcaagcaa a gaaatcgtg 27720
ttaactttta taacatctag ttgacactgt cagacccaaa acttaataaa a ttttcactt 27780
gtacataaca gctagctgaa actgtaattt aattttatat tccctcggtc a attctagct 27840
aaattagcga ttctgagcta agccttcatt tcaaaattaa caaaaaaat g caatgaaat 27900
tttcacttgt acataacagc tagctgaaac tgtaatttaa ttttatattc t ctcggtcaa 27960
ttctagctaa attagcgatt ctgagctaag acttcatttc aaaaataaca a aaaaaatga 28020
attgaaattt tcacttgtag ataacagcta gctgaaactg caatttaatt t tatattcct 28080
tcagtcaatt tcagctaaat tagcaatttt gagctaagtg ttgttgtttc t taaaacaat 28140
gcaaattttg atggtttttc gtgttcagtg aacaacaaa caaacaaaa a attctggta 28200
aataaccaca agctgaaact gtgagataat tttttagtga ccattgagtg a ctgctcata 28260
gacagtggct tggaaatag actagaatga ttatctctca tgataacata t tatacagag 28320
aagttgggaa gaatgtaggt cattgtaaag cgacagacag gtcgcattga t caagagaa 28380
tataagtcga actctttctg ttggttaact gagggccaat gttatttctt a ttagggaaa 28440
attaacattt aaggagcaaa ggattgcaaa caaaatgcga taagatataat g attatagta 28500
ttttatcttt tgtaagtgtt gccataattt cagtaacgaa aaaaataaca a ggcaatttt 28560
agatgttagg aaaatcgaat ttgtctgact agccaacgaa tgttctcaat t gaagttatt 28620

-continued

```

gttctttttt aagatgtttt catacaaatt agtcagtttt cgaagcttca g ccacactta 28680
tccgaattga gcaatttcaa aactattttt tgtaaaataa aatacatctc c gaaaattta 28740
catcgagttc ccaacaatac tgtatggata gaaaatacct accaactactg c acatgaaac 28800
gctctgaaaa taatcggaaa ggaaatgaga acctttttaa tataaaatga g cacaataag 28860
taatactaac tttattgaga aagaacataa ttgttatgag aatagttttt a aatgaggtg 28920
agaaacagaa tatccctgag aataagtga gatacttgaa aattttgtgaa a tagtaataa 28980
gtaaaatgtt ttcacattag tataaacaat gacagagtca cgaaaagta c gggaaacat 29040
atgaagttta taatacagtg cagtacagaa aaggtacaaa gtttacaaga a tacaattgt 29100
tttttaaaaa taattttttt ttgaaggctt aaggaatac gattaagag c tactttctt 29160
ccaatacgaa gttgaattta aaatttataa ggaaaaaagg aaaaaatta a aaagcatat 29220
gaaaaatcgg ggcgcatttt tagtgcaaaa aattagatgg catttatttt a tcccatcca 29280
tctgaatctt cactgtgtgt ggattttattg tcgcatctt gatcgatcat t gtatcatca 29340
gcttctcctt cttgattgat aagaagacct tgcagttttt ccgaaagttc c gaaatcttc 29400
aaatcctctt ctctcaatgc atcatgcato ttctgaattt cagcggatcg t tcgctattt 29460
tgaataagtt ccatcagaca ctcaattttg ctatottttt ccatgatttc t cttttatga 29520
tttgcaatct gttcttcttt tgattcacat tctctctttg aattggctga a ataaaagaa 29580
aatgcttaca gatgtgtgta aaaccttag aaaactttca caagcttacc t gtcaatact 29640
tcaaattgcc ccaataagtt gtgcttccac tcttcagttc gaagtttaag a tcttcaact 29700
gatgtattaa gcgtggcttt ttctgctga gtgtttgcaa gttgcatctc t aacgccatg 29760
acggtcagat tatgttgatc caaaatatga ctg 29793

```

We claim:

1. An isolated nucleic acid molecule which encodes a polypeptide which catalyzes release of ADP-ribose from an ADP ribose polymer, having a nucleotide sequence that is at least 80% identical to the nucleotide sequence set forth at SEQ ID NO: 1.
2. The isolated nucleic acid molecule of claim 1, which encodes a protein having a molecular weight of about 110 kilodaltons, as determined by SDS-PAGE.
3. The isolated nucleic acid molecule of claim 1, wherein said isolated nucleic acid molecule encodes a protein having the amino acid sequence set forth in SEQ ID NO: 2, 4 or 6.
4. The isolated nucleic acid molecule of claim 1, having the nucleotide sequence set forth in SEQ ID NO: 1, 3 or 5.
5. An expression vector comprising the isolated nucleic acid molecule of claim 1, operably linked to a promoter.
6. A cell transformed with the expression vector of claim 5.
7. The cell of claim 6, wherein said cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.
8. A cell transformed with the isolated nucleic acid molecule of claim 1.
9. The cell of claim 8, wherein said cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.
10. An isolated nucleic acid molecule which encodes a protein which catalyzes release of ADP-ribose from an ADP-ribose polymer, wherein said protein comprises the amino acid sequences set forth in SEQ ID NO: 11, 12, 13 and 14.
11. The isolated nucleic acid molecule of claim 10, the complementary sequence of which hybridizes, under stringent conditions to the nucleotide sequence set forth in SEQ ID NO: 1.
12. The isolated nucleic acid molecule of claim 10, comprising the nucleotide sequence set forth in SEQ ID NO: 1, 3 or 5.
13. The isolated nucleic acid molecule of claim 10, which encodes a protein having the amino acid sequence set forth in SEQ ID NO: 2, 4 or 6.
14. An expression vector comprising the isolated nucleic acid molecule of claim 10, operably linked to a promoter.
15. A cell transformed with the expression vector of claim 14.
16. The cell of claim 15, wherein said cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.
17. A cell transformed with the isolated nucleic acid molecule of claim 10.
18. The cell of claim 17, wherein said cell is a bacterial cell, a yeast cell, an insect cell, or a mammalian cell.

* * * * *